

Key words: VZV/DNA sequence/genes/homology

The Complete DNA Sequence of Varicella-Zoster Virus

By ANDREW J. DAVISON*† AND JAMES E. SCOTT
MRC Virology Unit, Institute of Virology, University of Glasgow, Church Street,
Glasgow G11 5JR, U.K.

(Accepted 23 April 1986)

SUMMARY

The entire DNA sequence of varicella-zoster virus (VZV) was determined using the M13-dideoxynucleotide technology. The genome is variable in size, but the sequence which was obtained comprises 124884 bp. Analysis of the sequence indicated that the genome contains 70 genes distributed about equally between the two DNA strands. The genes are organized compactly, but regions of overlap between protein-coding regions are not extensive. Many of the genes are arranged in 3'-coterminal families, and at least one is spliced. The discerned organization of VZV genes and that deduced for herpes simplex virus type 1 (HSV-1) from published transcript mapping data indicate that these two members of the *Alphaherpesvirinae* are very similar in gene layout. Comparisons of the predicted amino acid sequences of VZV proteins with those available for HSV-1 proteins generally suggest evolution from an ancestral genome, and allow the functions of several VZV genes to be deduced, although limited regions where the genomes differ in functional organization were also identified.

INTRODUCTION

Most people contract chickenpox as children, and many of these suffer later in life the painful symptoms of shingles. Both diseases are caused by the same herpesvirus, varicella-zoster virus (VZV): chickenpox as a result of generalized primary infection, and shingles as a consequence of reactivation of virus which has remained latent throughout the lifetime of the individual. There are no generally available measures for the effective prevention or cure of either disease. Despite such a motivation for studying this medically important virus, our knowledge of the molecular biology of VZV is rudimentary in comparison with that of the other four herpesviruses which infect humans: herpes simplex virus type 1 (HSV-1) and type 2 (HSV-2), human cytomegalovirus and Epstein-Barr virus (EBV). One major reason is that the problems encountered in obtaining sufficient quantities of viable cell-free virus *in vitro* have made the genetic analysis of VZV particularly difficult. Consequently, the most fruitful areas of VZV research have been those which avoid this obstacle. One such area has been the direct analysis of the genetic material of VZV.

The VZV DNA molecule was first shown to have a G + C content of 46% by Ludwig *et al.* (1972). In the years that followed, several VZV isolates were analysed using restriction endonucleases (Oakes *et al.*, 1977; Richards *et al.*, 1979; Zweerink *et al.*, 1981; Straus *et al.*, 1981, 1983; Martin *et al.*, 1982). Significant contributions were made by Dumas *et al.* (1980, 1981), who correctly determined the molecular weight of VZV DNA to be 80×10^6 , and reported the genome structure of VZV and the first three restriction endonuclease maps. These results were confirmed and extended by further structural studies and the derivation of additional maps and construction of libraries of cloned DNA fragments (Straus *et al.*, 1981, 1982; Ecker & Hyman, 1982; Gilden *et al.*, 1982; Davison & Scott, 1983; Mishra *et al.*, 1984). Additional structural features emerged from initial DNA sequencing studies (Davison, 1983,

* Present address: Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

1984; Davison & Scott, 1985). In summary, the VZV genome is a linear double-stranded DNA molecule consisting of two covalently joined segments, L and S. L comprises an unique sequence (U_L ; approx. 100000 bp) flanked by a small inverted repeat (TR_L and IR_L ; 88.5 bp). S contains an unique sequence (U_S ; 5232 bp) flanked by a large inverted repeat (TR_S and IR_S ; 7319.5 bp). The genome is not terminally redundant, and possesses an unpaired C residue at the 3' end of L and an unpaired G residue at the 3' end of S. Virion DNA contains two major and two minor genome arrangements differing in the relative orientations of the L and S segments; whereas one orientation of the S segment is present in 50% of virion DNA molecules and the other in the remaining 50%, one orientation of the L segment is present in approximately 95% of molecules and the other in only 5%. It has been reported that a small proportion of virions contains superhelical circular DNA molecules (Straus *et al.*, 1981; Kinchington *et al.*, 1985).

As a culmination of our own studies, the VZV DNA sequence is presented in this paper. Thus, VZV is the first member of the *Alphaherpesvirinae* whose genome has been completely characterized at this level. The usefulness of comparisons between VZV and HSV-1, also a member of the *Alphaherpesvirinae*, became apparent when Davison & Wilkie (1983) observed from DNA hybridization experiments that these viruses possess several conserved genes arranged colinearly in the genomes. The hypothesis resulting from this finding, that the two genomes have similar gene arrangements, was confirmed by comparison of the VZV gene layout deduced from the DNA sequence with that of HSV-1 proposed from currently available transcript mapping and DNA sequence data. Consequently, the functions of several VZV genes can be identified on the basis of our knowledge of the molecular genetics of HSV-1, which far exceeds that of any other herpesvirus.

METHODS

Recombinant plasmids. Plasmids comprising *Kpn*I or *Sst*I fragments of VZV DNA inserted into the *Pst*I site of vector pAT153 have been described previously (Davison & Scott, 1983). Additional plasmids consisting of *Hind*III or *Eco*RI fragments of VZV DNA inserted into the *Hind*III site of direct selection vector pTR262 (Roberts *et al.*, 1980) or the *Eco*RI site of vector pUC9 (Vieira & Messing, 1982), respectively, were characterized on the basis of published *Hind*III and *Eco*RI maps (Straus *et al.*, 1982; Ecker & Hyman, 1982; Mishra *et al.*, 1984). For S1 nuclease analysis of the mRNA encoding deoxypyrimidine kinase, VZV *Pst*I o was subcloned from the plasmid containing *Hind*III b plus l into the *Pst*I site of vector pUC8 (Vieira & Messing, 1982). All plasmids were propagated in *Escherichia coli* K12 strain DH1 (Hanahan, 1983).

DNA sequencing. DNA sequences were obtained using the M13-dideoxynucleotide technology (Sanger *et al.*, 1980). Plasmid DNA was sonicated and then precipitated using polyethylene glycol to give random fragments 400 to 1500 bp in size. The sheared ends of the fragments were repaired using T4 DNA polymerase in the presence of the four deoxyribonucleoside triphosphates. The fragments were then ligated into the replicative form of vector M13mp8 (Messing & Vieira, 1982) which had been linearized using *Sma*I and treated with bacterial alkaline phosphatase. Ligated DNA was transfected into *E. coli* K12 strain JM101 (Messing, 1979), and clones bearing inserts were identified as clear plaques in a bacterial lawn using isopropyl β -D-thiogalactopyranoside, an inducer of the *lac* operon, and 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside, a histochemical substrate for β -galactosidase. Recombinant bacteriophage DNA was prepared under conditions of good microbiological practice, and clones for sequencing were identified by hybridization of the appropriate nick-translated VZV restriction fragment. Clones were sequenced using pentadecamer primer (New England Biolabs), [α -³²P]dATP (PB.10204; Amersham) and the Klenow fragment of DNA polymerase I. The latter was obtained from Boehringer Mannheim in earlier stages of the work, but for the majority of sequencing the Klenow fragment was purified from the genetically engineered strain of *E. coli* described by Joyce & Grindley (1983). Sequencing products were separated in 0.35 mm 6% polyacrylamide-urea gels containing a buffer gradient (Biggin *et al.*, 1983). Each gel was bonded to one glass plate prior to electrophoresis (Garoff & Ansorge, 1981), and then dried prior to autoradiography.

Plasmids containing the following VZV DNA fragments were sequenced in their entirety: *Kpn*I t, *Kpn*I c, *Hind*III a, *Hind*III e, *Hind*III d, *Hind*III b plus l, *Kpn*I b, *Kpn*I i, *Kpn*I l, *Sst*I g and *Sst*I f. Junctions between fragments were established by sequencing specific restriction fragments from additional plasmids.

Data handling and analysis. DNA sequences from individual M13 clones were read using a Summagraphics digitizer pad with a program written by P. Taylor, and were compiled and analysed using the programs of Staden (1982) modified by P. Taylor for a DEC PDP-11/44 computer operating under the RSX11M system. Open reading frames were identified using the program of Blumenthal *et al.* (1982), and translated into amino acid sequences using a program devised by Taylor (1986). Codon usage was examined using the program of Staden & McLachlan

(1982). Sequence homologies were analysed using the matrix comparison of Pustell & Kafatos (1982) and the optimal alignment program of Taylor (1984). Hydrophobicity profiles were prepared using the parameters described by Kyte & Doolittle (1982). A search of the Protein Identification Resource (release 5) compiled by the National Biomedical Research Foundation (NBRF) for homologues of VZV proteins was carried out using the WORDSEARCH program of Devereux *et al.* (1984) in a DEC VAX11/750 computer operated by Edinburgh Regional Computing Centre.

S1 nuclease analysis of VZV deoxypyrimidine kinase mRNA. Cytoplasmic RNA was prepared by the method of Kumar & Lindberg (1972) from uninfected human foetal lung cells or from VZV-infected cells showing 50% c.p.e. Structural analysis of the deoxypyrimidine kinase mRNA was performed using the S1 nuclease digestion procedure of Berk & Sharp (1978) modified by using 5' or 3' end-labelled probes. DNA/RNA hybridization and S1 nuclease digestion were carried out as described by Rixon & Clements (1982), except that 15 µg RNA was hybridized at 45 °C to less than 1 µg end-labelled DNA fragment isolated from the *PstI* o plasmid. The digestion products were separated on DNA sequencing gels and detected by autoradiography.

RESULTS AND DISCUSSION

VZV genome size

The entire VZV DNA sequence is shown in Fig. 1. It was derived from approximately 1.2 × 10⁶ nucleotides of data, and about 97% of the sequence was determined on both strands. The genome size, from the plasmids analysed, is 124884 bp, and the G + C content is 46.02%, impressively close to the value of 46% derived from buoyant density centrifugation by Ludwig *et al.* (1972). The sizes and G + C contents of components of the genome are as follows: U_L, 104836 bp, 44.33% G + C; TR_L and IR_L, each 88.5 bp, 68.36% G + C; U_S, 5232 bp, 42.78% G + C; TR_S and IR_S, each 7319.5 bp, 59.04% G + C. The significantly higher G + C content of the inverted repeats is a general feature of herpesvirus genomes, and has been discussed previously for VZV (Davison & Scott, 1985).

Five regions of the genome contain tandem direct reiterations of short G + C-rich sequences. One, in TR_S, is a duplicate of that in IR_S, and so the four unique reiterations are denoted R1 to R4 in Fig. 1 and 2. Regions of the genome which vary in size between different virus isolates have been mapped by Straus *et al.* (1983), and correspond approximately to the locations of R2, R3 and R4. R4 (109762 to 109907 and 119990 to 120135 in Fig. 1) has the structure AAAAAX, where A is a 27 bp element and X is a partial copy of 11 bp of A. Casey *et al.* (1985) reported that the copy number of the 27 bp element varies between virus isolates.

The region containing R3 (41453 to 41519 in Fig. 1) is the most variable in size between virus isolates (Straus *et al.*, 1983). Moreover, fragments containing R3 are particularly difficult to clone in *E. coli*, and those cloned fragments which are obtained may differ significantly in size from the virion DNA fragment (Straus *et al.*, 1982). Thus, the R3-containing clone which was sequenced (*HindIII e*) is smaller by about 1000 bp than the estimated size of *HindIII e* cleaved from virion DNA. In this clone, R3 has the structure AAAAABAX, where A and B are unrelated 9 bp elements and X is a partial copy of 4 bp of A. Preliminary analysis of an independent clone of *HindIII e* which is about 500 bp larger showed that the additional sequence is contained within R3, and that the reiteration contains a complex arrangement of 9 bp elements, including one not present in the smaller clone (data not shown). The sequencing results, and the discrepancy in size between virion and cloned DNA fragments, imply that R3 may contain in excess of 100 copies of the 9 bp elements in virion DNA. Presumably, this highly repetitive structure is unstable in *E. coli*, so that stable clones are obtained rarely and lack many of the 9 bp elements.

In the Dumas strain, R2 (20692 to 21017 in Fig. 1) has the structure ABABAAAX, where A and B are 42 bp elements differing in a single base pair, and X is a partial copy of 32 bp of B. Again, variation in the copy number of the 42 bp elements results in size heterogeneity in this region of the genome in different isolates (P. Kinchington & J. Hay, personal communication). R1 (13937 to 14242 in Fig. 1) is a rather complex reiteration containing four elements: A, B, C and D. A and C are 18 bp in size and differ in a single base pair, B is 15 bp in size and unrelated to A or C, and D is 15 bp in size and consists of the first 6 (or 7) bp of A or C linked to the last 9 (or 8) bp of B. R1 also contains a partial copy (X) of 3 bp of A, B or D. The sequence of R1 was

TRL ---->---- UL

AGCCCAAGCCCCCTCGGGGCCCCCTCGAGAGAGAAAAAAAAGCACCACCTCCCCGGCTTGCAGGGGACCCATGGGGGGATGGATTGGGGAAACCCCCCCC
120
 AGCTTTAACAAAACCCGGCCCTTTCGGTCCACCCCTCGTTACTGCTGGATGGCGACCTGCACACTCCCGCCACCTGGGACCCGGTACGGTACGGTCCCC
240
 GGATGACGTTGGGACCCCCATCCCTACCTACCCACATAGCCGAGGGTGGCAGACGGGCCCCCTTACAGAAGCCCGAGAGTCTGGTTCTCCCGCTCT
360
 GGAGAATGGGACCCACCAACAGCTTACGGATGGCTGACTGCCCTTATGATGGAATCCACAGACTCAGCTGGCTTCTAAGAATTGCAATGCTGT
480
 TTTGGTATTCTCACCCCTACTGCTGCTGGTCCCATTTGGGGAGGAACCTCCAACTCAACTACATGAAACTACTGTCCGGAGGGGAAGGTATTT
600
 * - E R K
1
 120
 GCAGCTTGCTCCGGGATCTGGCTGGGACATCAACCCAGTGGATAAGAACCCGGGGGGTCTGGGATCCAACTGGGATCTGGGATCCAAAATA
720
 C S T A R T H Y V F A I Y T V L A L T A M Q L V V G L V N C L R D M Q N K L Q A
66
 GGACGCCCTTCTCCGGGATCTGGCTGGGACATCAACCCAGTGGATAAGAACCCGGGGGGTCTGGGATCCAACTGGGATCCAAAATA
840
 S A K R R S R P R P V D V L P I L V A P P R T O V V L P S A N N Q M L E T H Q Y
26
 CATAAACACAGAGTCTGATCGCTACAGATCCGGACACCTCCGGTACCCCTACTCGGATACCCCTGGACATTGGGATCCAAAATA
960
 M F V S D T D S D S E R V G E P V G Y E S V R S M
1
 GCTACAGCTTATATAAATTATGCGATACATCTTAAGTGCATCCGTACGTTTATACATTGGCTGACGTGAAAAGACTGTGTTACCAATAAGGT
1080
 2
 22
 TTGGGTTGTTTAAGCTTTATCGTAACCCACCCCGTAAATCATAAAATCGTAAATTCTGAGACACTTCGATATGGGATGTTCCGCAATTATGGGCT
1200
 V R P S L N A T A E E N P A S E T R C L L R V L A G R T V D L P G G G T L H I T
62
 TGCGTCCCAGTTAACCCGGGACCGGGTCTGGGATCTGGGATCTGGGATCTGGGATCTGGGATCTGGGATCTGGGATCTGGGATCTGGG
1320
 C T K T Y V I I G K Y S K P G E R L S L A R L I G R A M T P G G A R T F I L A
102
 GTACCAAACCTATGTAATTGGGAAATATGCAAAACCCGGGAAACGCTTACGGGCTTAATAGGGCTGAATAGGGCTGGGAGGACATTATGGG
1440
 M K E K R S T T L G Y E C G T G L H L L A P S N G T F L R T H G L S N R D L C L
142
 TGAAGGAAACCGATCCACACGGCTGGGATGATGTTGACGGGCTTACGGGATCTGGGATCTGGGATCTGGGATCTGGGATCTGGG
1560
 M R G N I Y D N H M Q R L M F W E N I A Q O N T T E T P C I T S T L T C N L T E D
182
 GGGGGGTAATTATGATGATGCAATGCAACGCTTACGGGATCTGGGATCTGGGATCTGGGATCTGGGATCTGGGATCTGGG
1680
 S G E A A L T T S D R P T L P T L T A Q G R P T V S N I R G I L K G S P R Q O P
222
 CTGGTCAAGGGCCTACCGCTAGGGGACCTCCCAACGGGAAAGCCAACAGTTCCAACTTCGTCAGGAGGATCCCAACAGGG
1800
 V C H R V R F A E P T E G V L M -
3
238
 TCTGTCAACGGGTTAGATTGCGAACCTACGGAGGGCTATTGATGTAATCACTAAATAAAACACCTTTTCTGATTGTAACGTT
1920
 - L G G
177
 ACAGCCGCTGGGTTTCCCGACATACAACATGATGCTATGCTCGGATGCCACGGTCAACACTCCGGGAGAAGGGGATTACA
2040
 V A A R A K G G C V V H D Y A E S A G T W C E A S F P S . K C H C H Y G I A A L H
137
 ACACCCAGCTGCCGACTCCAGCATCTGCTGAGTTGGCGCGTGAAGGGTGCATCGCATAGGGTTATAATTAGGGCTTCCG
2160
 V G L Q G S E L N N Q Q T A A S F P A D C L T H Y N A N E P L L R Q S N L L S C
97
 AACGGCTGAGGTCACACATACATTGGGATTCTGAGGGTTATCTGAGCTCCAAAGTCCAACTAAAAAGCGTGAATCAT
2280
 F P Q L D V Y M P S E S P K D R R G L G I L F A H L D D A R G C T A R L A C L R
57
 TTACACCGTACAGGGGGATGGCGCTGGCTCATGAGTTGGCAGGGCATGTCACCTGGGATATAACTTGTTAAACGAC
2400
 K V G Y L P S P T P A H S N A P C T N T T E L A L P P I V Q T F S A L P D P K L
17
 TTCACTCGGATGGGTTGACTGCTTCCGGCTTGTATCCATTAAATTAAACGCTGGTACAGCTGGTGTGTTTACCG
2520
 N V R I P Q S S E S A G T T D M
1
 GAAACCAATGTAATCCGCCCATATCCCAACGCTGGCTTAAACCTCCACAAATTACATTGTTTAAATGATACAC
2640
 TGGGTGATAATGATATAAAATTGATATGTTTACAAGCATGAAATAGGTATTACTTCAACGGTAATATGCTTAATGAT
2760
 CATGAAAAGTATTITTTATTTAGCTAAAGGTACTACCTAAATATTACCGTATGGGGGGCTAGAGAATGCCCC
2880
 4
421
 ACTCGGTTGGGGGGTATTAGGGCTCTAAACACGGGAGACATGACCCGGGTATATTCTGTAACACTGAGCTTACACT
3000
 E R Q H T I L A E L V G A L C S G P T Y E Q L V Q V H C S I D D Y E V F K S G R
381
 GGACAACTATTAAGCAATGGGGCAATCATAACAAACATAAGTAGTAAATACCGTATGACTCACTAAACCTG
3120
 V V I N A I R A I N V F N Y T T I C T I D S F R Q Q R L L E P L T R S G L D N D
341
 CCCTGTTAGAAGAGACATTTACAAAGGGCCAGCTTAACTTTAAATCTCCAAAAGTGACTCGAGGGTGC
3240
 R N L L L C R V F P G L K L K L H E L L S K S T A V I P N N T Y I P L N K A A L
301

ACATTTAACTCAGCTAACAGTCATCCCGAGACTCCAACGCTCAATCAAAGATTCTCACGATGACTCTCACGCACGCCGGCAATACGTGAGTCATTTATGACTCAA
 M K I N T L L E D A S E L A E I L S E G R I V R E R L A R A I R S D H K Y S E F 3360
 261
 AGGTACGATAAAGTCTATGTCGCTACAACATCAACTCCGGCAAGATGTGTTTGTTTATCCCCGAAACATCCACCGGAAGCCCATGATCACCCCTCTTATGTGGCATATCGGA
 T R Y L E H G Y L N H L E P W S T N Q K I G P F C G G S A W S D G E Q I T A Y R V 3480
 221
 CTACAGTTTCAATTGTTCATCTAAATGGCGTACCGAGTCATGGTACCGCTGGCTCCCGCGGAGACGACTTCATAGCACGCCGTAAATTGATGCCGATCCGGATATCAGTC
 V L K E I T E D O L H R V S D I T V S A G A T S V E I A R G T I R D Y P I D Y E 3600
 181
 CTTTCGAAATCGCTCTCCGGCGCTCTCTGGAAAATCGAACCTGTACGATTCGCTATGTCATCTTCTCCCGGTGTCATTGAGGGCGTTGAGGACGCCGTC
 K R I R E R R A D R K S F R L R Y S E D H T Q D N R E G T T M A P P T T P R R R 3720
 141
 TTGATTGACAGGATCGATCACGGTGTTCCTGAACTTGGTGTATAAGATCTGGATGATCGCAGTCGATGCCGATATCGACGTCCTTCAGTCCTCTCCAT
 R N S L S R D R H K R S S Q T H Y S R S S R R H G R E I R A Y G T E Y D R G G H 3840
 101
 GATGGTTGAATCGGGTAATACAACAAAGTTTCGGCGATTGTTGGTAGCTTCACGCCCTCCGCTGCTGTTGGAAATACCGTGGATTATATGCTGATCTGCACTACGCT
 H N S D P L V V V L T K P R N H H Y S E R R G H R R K S Y R P N Y A T D A T R E 3960
 61
 CCACATACAGCTTAGCAGTGTGGAGCTCCCTGGAGTGGAGCCAATAGCTCATCATTTGCCAACTGGTACTTCATGCCAAAGTCATGCCAAAGGTCGCTGGTACAAAT
 V Y V T R S T T S D E G P T S G I A E D N H A W D T V E L A F D D S P E D P L L N 4080
 21
 TCATAAAGTCTCACAAATAGTACAGCACGCTGGGTGGTGGAAATTGAGCAGAGGCCATGGCTGAAATATCTGACAATTGCTGTTTGAGTTCGCTGTATCTCCGCAATGTTG
 W F D E C I T S V D P D T P I S A S A M 4200
 1
 TAGAATTATAGGCTCACCAACCCCGAACGGCGTGTAGTCACATGATAATTGCTCTGGGAGTTTACCTTCCCACAAACAGTACCTGCACCCCTTGTCGTAATGCATAAA
 - H K Q S H E S E G F L S V Q V R Q E Y H M F 4320
 319
 ATAAACCACTGCTAGCAAATATGACGATAAAAAACATTATAGCAAGGCCGACATTACTGTAGCGCAACATGTTGTCATATACACCGTATTCCCCCGATTGATGATTAA
 I V V A I A F I V I Y F C K I Z A L G S M V T A C C T T C I G R I G G T N I H N L 4440
 279
 ATGATTATCTGGGGTTGGTTCATACAAGATAAGCTCTACTATAGCGAGGCTGTCATACAACACCCAGGCCAGAATCGGAATGTTGTCGATATAACCGCAGTGGT
 H N D K T P K P R V Y S I L E V I A L T C V V V V A L I R I Y T P Y L L A C P T 4560
 239
 ATATGCAACGCCAACCGTAAAGCACAATACCCAGATGATAATAGCGATAACCTCCAAACCATCAATAACGTAACACCTTATGTCATATAAAAACATTAGGGTCAGCATT
 Y A V G L T L L Y I C G S S I H A I V E L L H L L T V G K H M Y L F S I P D A H 4680
 199
 AAATACTTFACTCATACCATCCGTGCGATGGAAACATCACATAACACCTTGCAACTTGTATATGGGTAACCAAGAAGATGTCGAATAACCCGTTGTAATTCACTGAAATA
 F V K S M G D R R M S V D C L L R A L K T Y P Y G L L I N S I V R T V Y N L S Y 4800
 159
 TGATGTGGGGGATATTAACCTACAGGATGTCGAATGGCCAAACATACGCTGTTGAAATGTAACATACCATACATACACCCGAAACAAATCTTGTGAC
 S T P S I L E C S S R F P G F M R R I R R F Q L Y M G Y V F W A F F I M K L Q V 4920
 119
 GCACAAAATAAGCGTACAATACGTTGCTCCAGAACAAATTGCTGCAAGGGTGTAGAAATACGCTTGTACCATATGCTGATAATAGATTTCCGGCAGCT
 C W F Y A H C N R T G L V I R I K D H L P T S I A T K V M T D G S L L N E R C S 5040
 79
 GTAAATCGTATCCGAGATAGGCCATCCAAAACCGTTGAGTGGTTACAAACGTTACATATAAGAGAGTTGTTAAGACCCCGATACACCGTGCACCATTAACCGTGGTGCATA
 Y D Y G S L G D L F T S H N V F T V Y I L S N N Y S G N V V P G G N I V T T A Y 5160
 39
 CACACACTGTTCAAACCTTACAGGCGGTATACCATAGGGTAAACACGCTGTCGCTAAGTAGACACATAATTAAAATGTTCTGCTGATAACGCTGCAACCCG
 V C E H E F K Y R A T Y W L T F V A H G S L L C M I I F H E T K I G L A Q M 5280
 1
 TGGAGAGTGGCAATTCAAGCACGATGTTAGTACACCGGTTGGTTAATCTGAACTTAAATTGGATAATTAGGTTACTCTAGAGTAAGATTGATGTCATGCGATTGCTATGCACTT
 - S S T L H P Y N P V E L T F I T H N R N S D C K 5400
 1060
 TGAGCAAACATTGTTGTCGAACGAAATACACAAACGGTTGATGTCACCTCGCAGAGACAAATGTCGGGGAGCCGTTCTCCCGCGATGGGGATATGCAAGACAAGTGA
 T A F C Q Q A L S I C L R N H H D Y R L S V F T R P A T R G G R H P Y R L C T F 5520
 1020
 CCTTTTGTCCGCAATAGCGTGAATAACACCCAGTCCCTTGTGGCATAACTTGTGATGTTAGGTTATTCGCGATACGCCGGGGAAATGAAACGCAATATGCTCCAC
 G K T G C I L Q F L V W D R K I A I C K S S T L T Y E R D R G P F H V A I H E V 5640
 980
 AATAGATCTAAATTGTCGTCGACAAGGCCCTCAGTTGTAATGCGTCAGACAAAGCTTGTAAATTGCGCTTGTAAAGATATTGCGGGAGCTAAATATTGTT
 I S E L I T S D V F A E L T F A D L C T V G R E K L A K T L S I Q P S S F I Q K 5760
 940
 ATTACGGCAACCTTACGTTCAAACCTCTGCTGTTAGGTTATCCCGCCAGGTTATGTAATGCACTGGACATTGACTGGCTGTTGAATGAAATGAAAGTTGCGGGTTCT
 H R A V K R E F F E A Y E G G L N H L I F Q V P V N S Q P R S H F H F N A P N R 5880
 900
 ATGTGATGTCACAAACGCTAATATCAACTGCTCAGGTAACACATAAAATGGGAGTAGTTGTCACCCGGCTCCCTGTGGTTACTTGGAGAAAAAGCCAGTCTAAACT
 H S T Y F A L I D I C Q E P V V Y F P L L Q G V A T G T T T V K S F F P L R L S 6000
 860
 ATGTCGGTGGCTATAACACCGTATCTATAACGAAAGTCCCGTAAATACGGACCAATATTCACAAATACCGCTTCCAGCAACCCGCTGCTGTAATATTGTCAGTCAACCCCTT
 H G H S Y V V G T D I F S F D R L Y P G I Y E V F E R E L L V A Q O L I Q A F G K 6120
 820
 TAAAGTGGAGACCCCACTAACCGCATAGGGATTGGATTGGTAGCGCATACCTGAAACCTTATTCACAGTTACAGGTTAGAGTTCATGCAAGTTTCAATTGTTGATACATC
 L T S S G V L A Y P N P I P V C V R F G I K E K C N C P L T E H L N E N N S V D 6240
 780

GCGCTGTGATGGACTTCAGACGTTGTCGTGATCAAAAAACCATACATCCCTGTATAATTCTCTACACACGTYATAATTGCCATTTCATGAAATCGATGTCAGAATG
 A H T H V E S T T Q T D F F N Y D E T Y N E E V C T Y L E G N E I Y F D I D S H 6360
 740
 GCTGGTTATATCCAATAAAATCATCATCCAAACCTCAACGGTAGGTTCAAGGACATCGAGTTTATAAAAATAACATGGGCTTGTAGGGTTACACGCCCTTGGAAAAAGTA
 S T I D L L N O O D L V E V T P E P C A T K Y F Y C P D K N P N V V A K P F L L 6480
 700
 TTGATGGCCGTTAAATACCATGACGAAATGCTCGATGCCGATGTTAAACCCATGGATGGGTTCTTATATGAAAGTCACATCAAGTATGAGGTTGTGATTAAAGATT
 Q M A T L I G H R F A R M G A H L I G L P I P K R I H F D V D L I L N T I I L N 6600
 660
 TGTTAAATAGCTCATCCGTTATATAAAAGCTATCTGGGTATGTTGATGAAATTAGAACGTTTAAACAGCTAGATAATAGTAAAGTCACATCGATCTCGTAGTG
 T N F L E N R N I Y L Q D K P I N S S I K S V N K V S T S L L L T L Q M D R L S 6720
 620
 AGCGGAAACAAAATACATGGATTAATTTGTTAAGGCTCTCCGCAATTATCAGGCTCTGGCGYAAAGCTAACGTTGTTATTGATGACCCAGTATCAGGAA
 A A V F N C P N I Q K L D E A I L R A E H P L T Y R N T I S S W T D N A I V A F 6840
 580
 TGCTGGCCGCGTGAAGGCAAGGCTACCGGATATAACAGGCTATGGCCAGTACCTCGAGATGCCGATGGGCTCTAATGGAGTTTATAACTCAGGATGATCATGTC
 A Q A G H P L A Y R Y Y P M P G T V E S H G I L A E L P T K Y S L I S V D H T A 6960
 540
 TAGCCAGTGCAGCAAGAAAACAGTAATGTTGTAATCCCCGGCTCGTATCAAACACCAGTACGACCAACTTGGTAGGTGATCGTTTCAAGTGGCTGCTGTAACGCC
 I G T A A S F F L L L K T I G P S T D F G T R G S Q N P T D N A F H A A R L A E 7080
 500
 CGCGGAAACCCGAATCTCAAATTAGACAATTCTCGTAAACCCGGGTGGATTGAGGAATAGTGGAGGACCATCCATGGACTAAATTGTTTCAATGTTTCAATGTT
 A S V G S D E F M S L E D F G P P P N S P I T T S S W G Y P S F Q K E I N E V R R T 7200
 460
 TAGCGTTGAGCTAGGTCACATCGCTATAACTGCTAGGTTTGGCCGATCGTAAGACTTAAAGTATGTTGTTAATTGATATTTGTCATCTCAGGCTCAAGTCAGT
 L T T A L D C V G I F K S P K A A Y T L S L T Y T K T I T Y K H G I E P G L E T 7320
 420
 GACATCACAAATTAGCTCTTTTATATAGTCAGCATGTTGAGCAGAACGTCATGTTAAAAAAATTAGCAGTAGCTCTTCCAGGTTGGATGATTAAAGGACCGTT
 V D C I V N K K I Y D R M N L R S R V H N F F N A T A R K G L N S S K L L V P _ K 7440
 380
 ATTCACAAAATCTGAGTATGTAACCGCTGTAGGTTGCTGCTGATGTTCCGATTGAAACATTCAAATGTCAGATAAATATAACAAATTACGGCTCTGGAACTTAAGGCC
 N V F D S Y T V A Q L H D A I Q K R H F C E F H A L Y I Y D Y F E R D P V K L G 7560
 340
 TTTCTATCGTTGTAATATACCGATACTCGCTGTTCCGTTGCTGATGTTCTGTTGAGCTGATAGAGCATTGGCTGTCACACCTCTGTATATGTTGAGGA
 K R D N T I Y E S V A H I E T T D T H I R O L I Y S L A N K A T L G R T Y T S S 7680
 300
 ACAACAAAATGGAAAGTTTATCAAAGACAACAAAGCTCCGAAATATTGTAACCGCAACTAACATTAGGTAATGCCGACTTGGTAGTTAAACAAATCTTAACTGCTG
 C C F M S L K D F S L L D S I N Y G S C N P L A P V Q Y T L F L D K I A Q L D Y 7800
 260
 TAAGGGGTTTCCACGTTGTAACGTTGCTGCTGATGTTCCGTTTCAAGGCTTACAGGTTTCCAGGGCTCAAATCGAACAAATACCAAAACGGCAGCATCC
 L P T E L T N Y S T D T Q L L Y Y R T A L Y L T K E S P G F R V I Y W F P S C G 7920
 220
 ATACCCCGAGTAGCTGCTGCGATATGCGCCAATCTGACGTTGCTGATGGGCTATAATGTTGAGCTCTGACGACCAACGGATTTTAACTAACTTGCAGAGTGT
 Y G W Y L R R Y A A L V Q R E Y H A Y L S T L E Q R G V S K V L K C L T A E T 8040
 180
 GATGCGATAGGGCTGCTGCGATAATCCCTTCGGTTAAATGGTGTGTTGATCATGAGTTGTAACCTCCGAGTGAATGTCACACGTCCTCGATACATAGGGT
 I C L G N D S L G K R N L H H T T V M L T Q I V E S H I D F T E S I C L T D S I 8160
 140
 TATATGCCGTTAGGGGCTCCATACCATACCGCTTAAATAAAGCTTAAATACGTTGGGTTTAAACACAAACAAAAATAGGCCAGACCCGGGATCSTACATCC
 I H P N L P A G Y N L A K Y L A K F D T Q T K F C C F F I P W V R S R V D G T S 8280
 100
 AAATCCACCAATTAAATAAAAAAACGTTGACGTCCTACTACAAAATAATGCTTATTTGGTTCTCATGTTCTGTTACTTCAGTGGGCTTAGTTGGGATTACTTG
 F G G I L Y F I V N V D R S C F L H N I Q N E E D N E T V E R P R K T P I Y Q T 8400
 60
 GATCTCTCTCCCTCCATTGACAAAGACGTCATCAAGTCGGGAGTCCAAGTATAACTCACCACATACAGGGCTGCTGCTTACGCTGGGCTTAGTTGGGAGAT
 I E E R G N K V F V D D L D P T H T Y S V V Y L P E T S I O G T L L L L S H S I 8520
 20
 7
 TGACATCCCTTGCGAAATAACCGAATCGTGGTTGGAGGTTATCCATAGTCAACGTTGGAAACCCAGTCATGCAAGACGGTGTGTCAGCTTATGCGATAG
 A C G K T A F L L R I T P K S S K D M 8640
 1
 A R I P T E E P S Y E E V R V N T H P Q G A A L L R L Q E A L T A V N G L L P A 8760
 CTCGAATACCAACTGAAGAGCCATCTTATGAAGAGGCTGCGTAAACACGCCACCCCCAAGGAGCCCTGCTCCGCTTCAAGAGGCTTAAACGCTGTAATG
 P L T L E D V V A S A D N T R R L V R A Q A L A R T Y A A C S R N I E C L K Q H 91
 CTCTAACGTTAGAAAGCTGTCGCTCTGCAATAACCCGCTGTTGGCTGGCCCAAGCTTGGCCGACTTACGCTGATGTTCTGTAACATGATGTTAAACAGC
 H F T E D N P G L N A V V R S H N E N S K R L A D M C L A A I T H L Y L S V G A 91
 ATTACTGAGATAACCCGGCTTAAACGCTGCTGCGTCAACACATGGAAAATCAAAACGGCTTACGCTGATGTTAGCTGCAATTACCGATTATGTTGGG
 V D V T T D I V D Q T L R M T A E S E V V M S O V V L L E K T L G V V A K P Q 9000
 9000
 TGGATGTTACTACGGATATTGCGATCAAACCTGAGAATGACCGCTGAAAGTGAAGTGGCATGTCGATGTTCTTGGAGAAAATCTGGGCTGTTGCA
 A S F D V S H N H E L S I A K G E N V G L K T S P I K S E A T Q L S E I K P P L 9120
 CATCGTTGATGTTCCACAACCATGAAATTCTAGCTAAAGGGAAAATGTTGGGTTAAAACATCACCTTAAACCGAGGCGACACAATTCTGAAATTAA
 9240

VZV DNA sequence

1765

I E V S D N N T S H L T K K T Y P T E T L Q P V L T P K Q T Q D V Q R T T P A I	251
TAGAAGTATCGATAATAACACATCTAACCTAACAAAAACGATCCGACAGAAACTCTCAGCCGTGTGACCCAAAACAGCGAAGATGTAACCGACAACCCCCCGATCA	9360
K K S H V M L V -	259
AGAAATCCCATGTTATGCTGTATAAATATTGAAATAAAACTAAAAACGTTCTGGYGTATGTTTATTTGTATATAAAATAAACATTGCTGGCTGGCGYGGTTATTACATTAA	9480
TGTTTTAGAGAAATCGACATCGTTGTTCTTATCAGTTAACCCAATCCACCGCTTCCCCTGGCTGGGTGTCCTTATAGATCTAACGTTTAGATAACATTGTAACCC	9600
H K L L F D V D N T E K O T S G F G R T G R E S P T A I L D L T K T F Y N Q V G	357
GGATGCCACATTACCCGGATAGCAATAAGGAATGCAATTACTTAAACGTTGTTAAAGTGTATGGGTGTTGTGATCTTAAACAGGACCTGTGCAAGACGATCTCCCGT	9720
P I G C K G R I A Y P F A I N S L V N H K L H I Q T N H D I L L V Q A L R D G T	317
TTTATACCTATGTCATACCCGTGAGATTATACGTTAGAATTACGTTGTCCTCCAGCCATCGCGTGGACACAGCAATTACGCTGATCGCTTTGATCTTCAAATA	9840
K I R I D D G T L N Y V Y F K C H E G A P M A T P C V I I G S R S K S S R G F I	277
TAACCGTTTATACCGATGGTGAAGTCCCAGTCCTTAACTCGGATAACAAATTAAATTCCTTTTAAATAGGTTATGGTACACAAATCATATCCCGT	9960
Y A N I S P H Q L D W D R I I P L V I K I F E N R K L Y L N H Y P V C I D Y G A D	237
TCTTGGCGTTTGGATGATGATGTTGAGGTAAGGAAACATCGATATGGTATCTCGAGAACATCTGTAAGGTTGCCCCCTGCTGACCGTGGAAATATCGAAATTCAAGG	10080
E Q R K P H I I H K Y T L P V D I H Y E A S D R H L P Q G Q Q V T S I D A F E P	197
ATAACGGGTTTCTATAATTGACCGCAGTTGATAAGGGTGAACCTGATGTTAAATGGATCAGATGTTAAAGACGTTTTGGGAGAAGGGGACTTGTGCTTAATT	10200
I V P K E Y N S P S H S L P Q V Q I S K F I P D L H K L V N K P L L R S Q R L K	157
ACCGGGAAACAAGTAGATTGTTAAATGCGGGTAAATAACGGTTACTCCTGGCGGTAAACAAAAGGGCTGAAATTACTCCTGTAAACCCGATCAATAACTCGTTGGCGACAAA	10320
V P F L Y I T L H G P L I V T Y G P R Y Y L L A S I V G R Y G A D I V G N A V F	117
AAATTGTCCTCATCGCAAGGGCAGTATCTTGCATTAACAGTCGTTATCGGGAGGGCGCAGCTAACCAACAGCTCAACTGTCGATATAAAACGCCCGCTGTT	10440
F N D E D A L A T D K C Q I L L L A Y E N P P A S K Y L L E L Q Q M Y F G G R T	77
ACAGATTTTCAGATGGCAGTTGAGTTCTGTTCCGGAGTACAACGGTTGTCGACTTACTTTATCGCTAACACGGTTGAGCTGACATTCAAGGTTGAACTTCT	10560
V S K E S P L E L K K H N R L L L P Q H R S V K D D L V C Q L T O A C E P Q V E	37
ATTAACGGGTTTCTTTAAACACCGATTCGGATACTTGGGCTACAAAACATATCCTGTTAGGTTACTCCTGGCTTCAAGATGGCATTAACCGCTTCTGTTATTAACGATG	10680
I L I T D K L C R N P I T Q S C F M D G T H V A T E L I P D I V A E N N	1
CAAAATTATTTTGGAAGACAGCAGTGGGAGGCCAAACTTGCAGAACGGAATTGGCATGCGAGCTGTCGGCTC6TGGAGTTATATCGACGGATCAATGTCACCCCTT	10800
TTCTTCACCGATCCCTTTGGGGGTGTTGAGCCCTTATTCGTTAGCTTATCGCGTGTTCGGTTTTACTCGTTCTGTTACGCGACGTTGAAACCGCTTGGCTATA	10920
MASSDGDRLCR	11
AAAGGGGCAGGCGTGTATAAGAGGGCCCTGTTAACCGGCTGCGGTGTTGGATATTACGACCCATCGTTATACGTTATGGCATCTCCGACGGTGACAGACTTGTGCG	11040
S N A V R R K T T P S Y S G Q Y R T A R R S V V V G P P D D S D D S L G Y I T T	51
CTCTAATGCACTGGCTGTTAAACACCGCTAGTTACCGGACAATATCGAACCGCCGGCGAAGTGTGGCTGAGGACCCCGATGATTCAACGACTCGTGGGTTACATTACCC	11160
V G A D S P S P V Y A D L Y F E H K N T T P R V H Q P N D S S G S E D D F E D I	91
AGTTGGGGCGATCTCTTCTCCAGTACCGGGATCTTATTTGAACATAAAACGACCCCTCGCGTACATCAACAAACGACTCACCGGGATCGGAAGATGACTTGAAGAC	11280
D E V V A A F R E A R L R H E L V E D A V Y E N P L S V E K P S R S F T K N A A	131
CGATGAAGTAGTGGCCGCTTCGGAGGCCGTTGAGACATGAACCTGGTGAAGATGCTGTATATGAAACCCGCTAAGTGTAGAAAAACCATCTAGATCTTTACTAAAAATGGCCG	11400
V K P K L E D S P K R A P P G A G A I A S G R P I S F S T A P K T A T S S N C G	171
GCTTAACCTAAATAGGAGTCACCGAACGGAGCTCCCGGGAGCAGCGCAATTGCCAGGGAGACCAATTCCCTCAGCAGCACCAACGCCAACAGCTGCTGGCGG	11520
P T P S Y N K R V F C E A V R R V A A M Q A Q K A A E A A M H S N P P R N N A E	211
TCCCTACGCCATATAACAAACCGCTTGTGAAGCGCTGGCGCGTAGCCGCGCATGAGCACAAAAGGCTGCCGAAGCGCTTGGAAATAGTAATCCCGAACGGAATAACGCCG	11640
L D R L L T G A V I R I T V H E G L N L I Q A A N E A D L G E G A S V S K R G H	251
ATTAGACCGTTGTTAACCGGAGCGTATTACGGCTGATGGGTTAAATTAACAGCGCTAATGAGCAGACCTAGGTGAAGGGCATCGGTATCCAAACGTTGAGA	11760
N R K T G D L Q G G G W N G H E P M Y A Q V R K P K S R T D T Q T T G R I T N R S R	291
TAATCGAAAAACTGGAGATTACAGGGGGCATGGTAATGACCTATGACGCAAGTCGTAAGCCTAACAGCGATACAAACGACTGGCGTATAACTAATCGAAGTAG	11880
A R S A S R T D T R K -	302
GGCCCGTTCTGCACTAACAGAAACTGATACCGGAAATAAGGGATATAATACCGCAAGTAACGGTTACCCGGTTATGTTATGATAATAAAACGTTAAAGAGACGTCGTTGTTGTT	12000
ATAATGTCATATATGTCACATATAAACTGTTAAATGACCGACGTTATGAACTGTTATAATCGTGTACCCACGTTAGGCTAGGCTTGGGTTGTTGTT	12120
M E C N L G T E H P S T D T W N R S K T E Q A V V D A	27
GGAAATCGCTTATTAACAAAGATTTACTCTATAAGTGTGAGCTGAGCTGAGAACACGCAATCTAGTACAGATACGTTGAACTGAGTAAACCGAACACGCCGTTGTTGCG	12240
F D E S L F G D V A S D I G F E T S L Y S H A V K T A P S P P W V A S P K I L Y	67
TTGATGAATCGTGTGTTGGTGTGAGCTGCGATATGGATTGAAACGTCGTTATTCACATGCACTGAGTAAACCGAACACGCCGTTGTTGCG	12360

F T P T E R E N L A S F F T L Y V T L G G G H L N W I C R A T A M Y L A A P Y H 590
 TTACCCCAACGGAAACGGGAATGTTAGCGTCATTTTACGTGTATGTTACTCTGGTGGAGGAATGTTGAACTGGATCTGAGACCACTGCAATGTATTAGCTGCCCTAACATT 15360
 S R S A V I A V C E S L P Y Y Y I P V N S D L L C D L E V L L L G E V D L P T V 630
 CCCGTCGGCTTACATCGCGCTGTGAATCTCGCCCTATTACTATACCGGTTAATAGTGACCTTGTGATTAGAGGTTACTGTAGCGAGGTGACCTCCAACTGTT 15480
 C E S Y A T I A H E L T G Y E A V R T A A T H F M I E F A D C Y K E S E T D L M 670
 GTGAATCCTACGCAACTATTGCAACACGAAATTACCGGATATGAGGCTGTCGCACAGCAGCCACAATTATGATAGGTTGCGATGTTATAAGGAAAGTGAGACCGATTTAATGG 15600
 V S A Y L G A V L L L Q R V L G H A N L L L L S G A A L Y G G C S I Y I P R 710
 TAACGGCGTACCTGGGGGGCGTTTATGTTACACGGGTTGGGTATGCAAATCTCTCTCGCTGCTGCTGTTACCGAGGATGTTCAATTACATCCCCGAG 15720
 G I L D A Y N T L N L A A S P L Y A H Q T L T S F W K D R D D A M Q T L G I R P 750
 G T A T T T A G T G C A T A T A A T A C T T A T G T G C G C A G G C A A G C A G G A T G T C A G G C A T C A C T T A C T C T T G G A A A G A C C G G A T G A T G C A A A C T T T G G G G A T T C G A C C G A 15840
 T T D V L P K E Q D R I V Q A S P I E M N F R F V G L E T I Y P R E Q P I P S V 790
 CAACGGGACGTTTACCCAAGAGCAAGACAGGATAGTTCAGGCACTACCTATAGAGTGAACCTCCGTTTGTGGGATTGGAGGATCTATCCCAGAACAGCCCATTCCCTCGG 15960
 D L A E N L M O Y R N E I L G L O W K S V A M H L L R K Y - 819
 ACCTAGCGAAAATCTTATGCAATACAGGAATGAAATTCTGGGTTGGATTGGAAAAGCGTAGCCATGCAATTACTACGAAAATTTAAGGTTGTGATTITTCATTAGGATGAAAAG 16080
 AACGTTTCTAGCCACACCCACAAAGGAGTTGTAAGGAAATTCTCTTTAGACCTTAAATTGTTGTTGTTGTTGTTGGGGGGYCCGTGAGGATEGACCTTACAAGATATAATT 16200
 12 M F S R F A R S F S S D D R T R K S Y D G S Y Q O S F N A G E R D L P T P 36
 TGCCCATATCGCAATGTTTCTCGGTTTCGCGCTCTTCCAGCTGATGAGAACCGCTAAATCTTATGTTGTTACCAAGGTTATGCGGGAACGTGATTGCCCACACC 16320
 T R O W C S I S O R I T S E R V R D G C L I P T P G E A L E T A V K A L S E K T 76
 TACCCGGGACTGGTCTTATTCACCGATAACCGGAGCCGCTGAGGATGGATGTCCTTATCCAAACGCCCCGGAGGCTTGGAGACGGGGTAAAGGCTTATCTGAAAAGAC 16440
 D S L T S P V L Q S T E R H S V L L G L H H H N N V P E S L V V S C H M S N D V H D 116
 CGCACGCTTAACATGCCGGTTTACAAAGTACCGGAAAGCACACGTTCTGCTGTTGATTACACCATATAATGTTCTGAACTGTTGCTCTGTTATGCTAACGATGTCATGA 16560
 G F N Q R Y N E T I Q R C L D D L K L S G D G L W W V Y E N T Y M O Y L K Y T T 156
 CGGGTTTATGCGCTTATGAAACAACTTCAAGATGTTGGATGACCTGAAACTTCTGGGGATGGACTTGGGGTTATGAAAATACATTTGGCAGTATCTCAAAACACAC 16680
 G A E V P V T S E K V N K K S K S T V L L F S S V Y A N K P I S R H P F K S V 196
 AGGAGCCGAGGTTACCGGTACTTCAGAGAAGGTTAAATAAAAGCTAAATCAGGTTTGTGTTACCGTAGTGTGCAAAAGTATATGTTATGCTCCAGATGTCCTAACAC 16800
 I N S D Y R G I C Q E L R A L G A V Q K Y H Y F M R P D D P T N P S P D T R I 236
 TATAATTCCGATTACCCGGGAAATGTCAGGGACTACGTGAGGCTTACGTGAGGACTGTCAGGAAAGTATATGTTATGCTCCAGATGTCCTAACACCCAGGGATAAAGAAT 16920
 R V Q E I A A Y T A T G Y G W M L W F L D O V V D A R V C R H L K L Q F R R I R G 276
 ACCTGTACAAGAAATTCCGGCTTACACGGCTACGGTGGATGTTATGCTCTGGGTTGCTGAGGCTGTCAGGCTTACGTGAGGCTGTCCTAACACCTCAATTTCGAGGATTGAGG 17040
 P R A S V I P D D L L R R H L K T G P A V S A G T G V A F I L A A T T A S A L T 316
 GCGCCGCCGCTCTGTTATCCAGATGTTGCTTAGACGACATTAAACGGGCTCTGCGCTCAGGGGGCACAGGAGTTGCTTATGAGCAACACTGCAAGCGCTTAC 17160
 A L L R I S V L W R K E E W R D G L N G T A A A I V A A V E L I T L L H H H F Q 356
 TCCGCTTTCGCTTATGCTTATGCGAAAGGAAGAGTGGCGGGATGGTTAAATGGAACCCGAGCTGCAATTGTTGCGGGCTGAACTTATACGCTTTCGACCACTTCA 17280
 Y L I N N M L I G Y A C V M G D G G L N H D P Y I L K A L R A Q G R F L Y F A G Q L 396
 ACTTAAATATGATGCTTATGGATATGCTGTTGGGGGATTAACGATCTTATATAAAGGGCTACGTGCCCCAGGGACGGTTTATTTGGCCTGAGCT 17400
 V R T M S T H S W V V L E T S T H N W F S R A V A Q S I L A H G G K P T K Y Y A 436
 GGTCAGAACAAATGTCACACACAGTGGTTGTTAGACGACCCATGTTGGTTTCTGGGGCTGGGGATGGCTGAGGAGTATTTAGCAGATGGGGTAACCCACAAAGTATTATGC 17520
 Q V L A A S K R Y T P L H L R R I S E P S S V S D Q P Y I R F N R L G S P I G T 476
 TCAAGGTTCTGCGCCAGTAAACGGTATACCTGGTACATTTAACGCTGATATCCGACCATCGAGTGTCTGATCAGCGTATATGCTTAACTGACTGGGATCTCAATAGGG 17640
 G I G N L E C V C L T G N Y L S D D Y N A S S H V I N T E A P L N S I A P D T N 516
 AGGTATGGGAATTGGAATGTCGTGTTACGGAAATTATTTATCTGACGAGCTAACTGCAAGTTCGCACTGTAATTACAGAACCGCTTAAACGATAGCACCCGATAACAA 17760
 R Q R T S R V L V R P D T G L D V T V R K H H C L D I G H T D G S P V D P T Y P 556
 TAGACAGCGGACTCTCGCGTTAGTTCGCTGAGACAGGGTTGGATGTAACCTGTCGAAAAAAACCACTGTCGACATAGGCCATACGGACGGTAGTCAGTTGACCCACGATCC 17880
 D H Y T R I K A E Y E G P V R D E S H T M F D Q R S O L R H I E T Q A S L N D H 596
 TGATCATTACACCGGATAAGGGCGGAAATGAAAGGTCGGGTTGGGATGAATCAAACACAACTGTTGACCAAGATGGCTACAGCTACATGAGAAAGCCACGATCTTAAATGATCA 18000
 V Y E N I P P K E V G F N S S S D L D V D S L N G Y T S G D M H T D D D L S P D 636
 CGTATATGAAAATACACCCAAAGGAAGTGGTTTAACTCATCTTCAGACCTGGATGTTGAGACCTTAACGGGATACACCTCCGGAGACATGCTACAGACGATGACTTACCCAGA 18120
 F I P H D V P V R C K T T V T F R K N T P K S H H - 661
 TTTTACCCAAACGAGCTCCGTTAGATGTAACACACGGTTACGGTTAGGAAAAATACGGCTAAGAGTCATTAAGTACAGCGGTTAATAGATAGTATGGACTAGGCACCTTGG 18240

VZV DNA sequence

1769

A V D V N N Y M Y T L L E R L Y P V G K R E N L H G P S V T I H C L G V L L R L 71
 GGC GG TAG AC GT AT GG A AC GT CA TG TA CAC AT TT G T G G A C G T T A T C C C T G G G T A A C C G C G A A T T A C A C G G C A C T G T A A C G T A C T T G T G G A C T T G C G G C T 24360

L T Q R S Y Y P I F V L E R C T D G P L S R G A K A I M S R A M N H D E R G T S 111
 ATTAACACAACCGTCACTATCCGATATTGTATTTGAAACGTGTACAGCAGGCCATTATCAGTGGAGGCCAAGGCAATTATGTCAGGGCCATGAAACCGATGAAAGGGAACTC 24480

D L T R V L L S S N T S C S I K Y N K T S E T Y O S V F R N S S T S C I P S E E 151
 GGACTTAACCCGTGTTCTACTATCATCCAACACATCATGTCTYCTATCAAGTATAAACAAACATCGGAACATATGACAGTGTTGCAAAACTCTCCACAGGTYGTATTCCTACCGAAGA 24600

H K S Q D N F L D G C P R Q T D K T I C L R D Q N V C S L T S T W P S R G H P N 191
 AAACAATCCCAGGATATGTTTGGACGTTGTCACGACAACACTGACAAGACGATCTGCGCCGACCAAAACGATGCACTGCTTACATGCCATCCGAGGACATCTCAA 24720

H R L Y H K L C A S L I R W M G Y A Y V E A Y D I E A D E A C A N H L F H T R T V 231
 CCATCGATTATATCACAATTTGTTGCAAGTCTTATTAGATGGGATGGGATGCAACATCGTGGAGCGGTTGACATTGAGGGGAGGCGATGTCGAACACTTATTCATACCGTCAAGT 24840

A L V Y Y T D T D L L F M D C I L L D A I P M F A P V V R C R D L L Q Y L G I 271
 GGCTTGGTTATACGACAGATCTGTTACTCTTCATGGCTGTGATATTGTTAGTCAATTCCTATGTTGCTCCAGTAGCTGCGATTGCTTCATATTAGGAAT 24960

T Y P E F L V A F V R C Q T D L H T S D N L K S V Q Q V I Q D T G L K V P H Q M 311
 TACATACCCCTGAATTGTTGGCCCTTGTGCTGTCAGCCGATTGTCACAGAATCTGTTCAAGTGCACAACTTAAATCTGTTCAAGTATTACCGATACGGGCTGAAAGTCCACATCAAAT 25080

D T S T R S P T Y D S W R H G E V F K S L T V A T S G K T E N G V S V S K Y A S 351
 GGACACTTCAACGGCTCCCCACTTACGACTCTGAGGACATGGCAGGGTTTCAAAGTCTTACCGTAGCGCACGTCGGTAAACAGAAAACGAGATGTCGTTCCAAATATGCACTC 25200

N R S E V T V D A S W A L H N L L P P S S P L D N L E R A F V E H I I A V V T P 391
 TAACCGATCGGAGGTGACAGTAGACCCACTTGGCTTAAACCTCTGCCACCCCTACCTCCCTGGATAATTGGAAACGCGATTGTCGAATATTCATCCGTTGTAACATATAATGCCGTGGTAACCTC 25320

L T R G R L K L M K R V N I N Q N T A D P Y M V I N T L Y H N L K G E K M A R Q 431
 ATTGACCGCGGCTGCCCTAAAGTTATGAAACGTTGTAATATTATGCAAAATACGGCAGACCCATATATGTTAAACACCTTATCATACCTAAAGGGGAAAAATGCGCTGCCA 25440

Y A R I F K O F I P T P L P L N T V L T K Y W H - * 455
 ATACCGACGTATTAAACAGTTTACCTCACTCCACTAAACACTGTTAAACAAATTTGGAAATTAAACACATAAGAGCAGCTTAATGTTTATTTGCT 25560

CGTATATACATGTTATAATCGTTTACTGTGCGGCGATAAGATGACTGTGTCCTCAAAAGTTGTTTATCTGCAATCATAATGCAAGTGGAAAGTCCGAATCGGGAGG 25680
 18 - L D H I V T G A Y S T S H R E F F N T H K D A I M F A L P F D S D P P 272

TGGGGTGTAAATAGTTGGTACATTAATCGTGTAAAGGCTGTCGGCGTGAATTTCACGTTGTTGTAATTGCACTGCGTCAACAAACGGGTTTGGGTCATGGGATTITAA 25800
 P T H F L K P V N I A S L L R D A S F K V Y Q T I A D V N V L R T K P A H S K L 232

AAACGCACACTCGATTCAACGGCTCCGAAAACAGTTGATTTCTGGTGTAGCGGGTTTCGGGTACATAGTTATGTTATACACGATGCGCTGGTATGCGCTCATC 25920
 F A C E I E V A E S F L O H I R T I A P K E P V Y N N Y I C C S A S T H I A E D 192

TGGGTTATAAGGCGTAAATTGACAAGTTACAACAAATAGTCGGTTATGGCAAAATGCAATAGCCGGAACGATGATAACAAAAAAATGCCCTATAAGAATCATAGTATA 26040
 R S I L D N F Q C T V V F L G N N R L Y A I A A F S S V F F I G E I L I N L I Y 152

TTTTCTGCAACGGATGGGTTGCGTACCTTCTTCAACATTGTACTTTTGTGATGACGGGTTTCGGGTACATAGTTATGACATTACGTTGACCTTACCGCAACGATTCCTCT 26160
 K E A V S P N D R V K E E L W Q V K Q Q I S P N H I T V N V Y Q V R L S E D G R 112

GAAACACATTAGTGAATTGACTATAACACCGCGCTGGACAAACCTCGACTCTGTTCACTGTAATGGTAATTCCTTGGAAAAGAGTTGGTTAGAGAGCCCAAATT 26280
 F L M L Q I Q S Y V R A H Y V E I C E Q E I Y Y H I D K Q S F L Q T L S G L N 72

AACATTACAGATCATCTGCCGCGATAAAATGAAAAAAATCTGTAAGTATTAGTCATCTCCGTTAAACAGTCAAGTATTGATAATCATCTCAATGATAAAATGCCCTTC 26400
 V N Y L D D A A S L F T F I F R Y F I L E D E T L C D L Y Q Y D D E I I F D S E 32

TAACCAACGATTGCAAGTCTAGGGACAGTAAATGTTATATCTGGACACTCCGGCTGTAAAAAAATGACTGCAATCTCTGATCCATTGGAATAGTTCCGTGTAATTAA 26520
 L W R H S I S L A R L N N I D P C E P R Y F F H S C D K Q D M 1.

TAAGGACAACACTGGTACAGGTTAATTCGCTCCGAAACAGTCGCTTTCGAGCTTACGAATTTCAGTAGTACATACCGTTTAAGGCCGCTTATAGGACGATAAGCAA 26640
 19 L A C S T C T L E G G A F L G S N T A K R I K C Y Y M G T K L G A K Y A R I L L 736

ATTCACTTATGGGAGGGCGGAAATTGTCGGCTCTGGCGTTCTCAATAAAATGCACTGTTGCAATAATGGCCCTTCTGACACATATCAACGAGATCTCTG 26760
 H M I K S A P I T G D P R E E I F L T M S Q S Q D I F P A R E A C M D V L D E Q 696

CTCATATTCAACGGCTTTTATTTAAAGAGTGGGTACTATTAGATAAACAGGCAAACGAGTATTACTGACCTTGGTTTCTCAAGTAGTTATAACTTCAGTGGTTTC 26880
 E Y E F A T K Y K L L P H S N S L C G F S R I V S K O N K E L I N I V E L R E 656

TTGACATGAATCATCTCTAGTCGCTCATAGGCTTAAGTCACTGGGTCAAGTAACTCACCGAGGTGGTACCTTCAAGTAGTTATAACTTCAGTGGTTTC 27000
 E C S Y N D R L E D M L D L N P R L L E G S T T V K S F M N N Y I P S F G E S C 616

CTCCGTTACCTGTCAGATGAAACTGTCGGCATTAACCGTAAAGAACCTGGACTTGTATAACCCATAAGGCCAAATATCATCGCAGGGTACACCATGGTAAACTCAATAACTTCG 27120
 E T V Q A S S V T P M L A F L Q S N Y L G Y A C I D D R L T C W P L D L Y S I T 576

AGAAAACCCATCTGGTAACTCCCTAGCATATTACTTCGGTAAACCCCTTAACGGGGCTAAGGCCAATCTTACATCTCCATGCTGTTTCATGCTCATACACAT 27240
 S F G D O H L W G K A Y K S E T F G K F P A L G G I K C M E N S T K N T E Y L M 536

TAACTCCGCTATTTGACATTAAACCGCTAGCTGGTGGGAAGTTAAATCAAATCCTAACGGGAGACAAGTTGATGTAACCCCTGATGCCAATGCCAAGTGATC60TTGTTTTAC 27360
 LEAIQVNLRBAPQSTLDLDFGLRLCTTHLGQIGIGLRSRNNKV 496
 ACCTTTACATGATTTTACATGGAAAGTCCACGGCCAGGACCCGTTAAAAAAAATAACAGTCGTTCTGCTGTCAATTGAGCGTGTAAATTAAATGACACTGGGCTTGG 27480
 GKCSKKCPFNGAALVGNLFIIVTTRATLQLDNLNFSVPGKS 456
 TAAGCACCGTTGAGATTATGCTGGCAAGATTACACCGCCAGTTGATGAGCGCTGCCTTGAACAAATTGACACAAATTGACCCGTGATGCAATTCCCTGGTATTICAT 27600
 LCTTLNISALNCVGHQHADAKOVIETCLNSGTIAHNGQTNM 416
 ATGATAATTACGATTACAGGCATCTTGAAACATTTAGGGGCTTCTGTTACACCGCACTGCGTATGTTGATGCGATATCTGAAATGGGAAAGAAACGCTAACCTTC 27720
 HYNRNCADKFMFLPSGTVAASRIITFAIDQIPVSSVGLGE 376
 TCTCTAAACGTTAGGTGAAGTGAATGCCCTCCCGTGTATGTCGAGGATACGGCTCTGTTATCAAAAGGTCGACTAACATTAGCCCTTTAGTAGCTTAGGTTA 27840
 RELBLYTSFTFAEGHLTRLIDARNDFLTNQVHSAGKLYSLY 336
 TCTTCAAAATTAATCTGGGTCCATAACACAAAATATGTTACATCGAAATTTICATCACGAAACATCCACGTGTGGCCAACAGTTGATGACGTCATGG 27960
 REFFLDPTWLCCFINDCRFIEDRYVLNGRTALVTQLDVHNP 296
 TTCTATGAAACACAAACCTCAGTTGGTGTCAACATCACTGTTAATGCCATAACCATGCAATCTAAAGTTTAAACTGCAAGAAGACCTTCGTTGATTTCCGTAGGTTAA 28080
 EIVVCVGTPRECDNSNIAMVNC DLLKLVALLGKTKNETPIL 256
 ATTTCAGACTCTGAGAGAAATTCCACCTCAGCTTGTAAACGGTCCACATGCCGTGATAGCTCGAACAGCTCTCCAAACAGTGTGGATTCGGTCAATTAAACA 28200
 NLSQLS16VGGGRSOLVLTGVDGTYIARVARYGVTISEPDNLYC 216
 ACTGGCGGTGCCCCGGTCTCGACCTTAAACATCATAACCCGGTGTAGCCGGGACAATTCTGACATGCCAACCGCTGTGAAAATACCGACAGACATCAGTCATGATAACCATC 28320
 SATAGTERGLFMMVPTAPVVKQCALATFFVRCVDTNTYGD 176
 ATTATTCGGGAAATAAGGTTGGTATTAGGCGGGTACCTGTTGTCACGGTGGCCGCACTGTTAAAGAATTGGCAAGCGACTCTAATTACCTCTCTAACTTAGT 28440
 HIGPILTAIKPLNVIETTYTAALRLFFFOCLSELKGEEELKT 136
 TAAATAAAAGTCCTGACTTTAAAGCAGACTGTGCAAGGGTAGCTAAAGGGGGTCAACCGAGTTCTGTCACGGTGGCCGCACTGTTAAACGGCTCAATAGCCGACGAAATTGCTCCCTGGCTTAAT 28560
 LYFDEYKLAQLGLTALAPYQDKLFPETIARRIEDRRGEI 96
 TGCTTGGCGGACTCGGGGAGTTAAACAGAAATTGGGGAGTCACCGAGTTCTGTCACGGTGGCCGCACTGTTAAACGGCAATGGAAAGTCTCCACACATGGTACACTCGCTC 28680
 AQRVRPTECLIPSTLWTENSYSRNLIRCHILEGCCCRYVRE 56
 ATCTTGCGCGTACCCGCTTAAGTTGAGACGATAGTCCTAAATACTCCATTAACTCCACCGGTGTGGTGTGATGTTAGCTTGTAGCCATGTTGACATAATCG 28800
 DQRTVAKLKSVITSIYENLEVPTTSEPPPIIYEKYGHQCLR 16
 GTTTATAATGTCATGACCGTATAAAATTCTTGTGACTCCATAACGGATAACGTATTAGGCTCCGGAAATAACCTTAAACCTTAAACTCACAGCTGAGTTAGCTCACATATTG 28920
 NIIDHVTNFIRKFEM 1
 TAGACTCCCTTATATGTTACGTACGGCTGCCCTCCCCAGTATATAATACGGCAAACCCACGCTATGTTAAATTCACTACATGCTTAATAAACATTG 29040
 -YYCEN 479
 20
 TCCATGTTATGTTACCCCCCACACACCCCTCTAACAAATAGTGGCACGTATAACCTCGAACCGTCCATGCCCTGTTGATATAACGCAAGACTCTGATGGAATTGTTCCAATTA 29160
 WTNTGGWVVGELWITPVNVYGGFREMRTYRVSESPITGIL 439
 ACATATGCCCCATACAGCAGGATAATTGTTGGGAGTCACCGCTCCGAAATCGCCGGTCTGATACAATGCTGTAGCCAAGTCCAAATTACTCTGTAATTGCGCAACTAC 29280
 TYAAAYMCSDLQTPLGRFDGTVQYLRLQBALNNNNYGTIE6ISC 399
 ATCGAGGGCTTGGGTCTGGATAACTGCAACGGGACGCCCTGGTTATATGGCTGGGGATTGCAACCCCTGAGTCCCAGGGGCAATTGCTGTAGCTGTC 29400
 RPSTPDNSLQVLPALARTHYPQHTNAVGETGWAANEARIRG 359
 CTAACAGGAAAGCCAATACACGACCAAAACATTGCTGTTAGTGGCTTAATGCAAGAATGCACTCACCTCTGTTGGCTGGCTGTGATATAACAAAACCAAGAACATGCT 29520
 LLFGIGRGFCQENLQLSALHLRVRGERQRRQTYVYFVHK 319
 TCAGTCGGTCCCGGGAAAGTGAATACTTGTCAACGCTCCAAAATCGCAGGGGGAGTGGCTGTGACCCCTGAGTGAAGTTGCTGACGTGCACTCTGGGACCTT 29640
 LGDASLHLDKDVDMFVCAPINATYRRTSTKIDTRAAEQPGK 279
 TGGCACGGGGTTATTCGATAAAATTCTGATGGTATTCAGCAGGGACCCAAAAAAACTTATAATCGATGTTGAAACATGACATTGACCATCGCACGTAACACTCG 29760
 AVATINCLNDQITYELLSS6FFSIFRHPFVHCQVWALCES 239
 AAAACCTTATGACCGCTTCCATACGACTGCACTCCATAGGCAGAACAAATTGCTGTTGTCAGGCTCCGCTGTATACCGTATATTCTGCGGCAATGCGGCAATGAAAC 29880
 FRIILRTEWRSCGYSVIACTRHNADAAQKDTYEEARCAAFIS 199
 TTAAATGACGTTACATGCTCTAACGGGAGTCATGGTAACGGTCAACGGTCAAGGAGGGCTGAGGATTGAGCTGATGCTGTTGTCAGGCTGAGTGAATGGTTGCTAGT 30000
 LSTVYARLGGGEGGVTPATCANFLNRLGELLYDPORGQAQT 159
 TAAGTGTTAGCTCTCCAACTAGTCAGATGGGAGGATTAAGCAGGCTGAGCCGAGCTGCTCAATTGCTGTTGTCAGGCTGAGTGAATGGTTGCTAGTACATGGGGTCACTG 30120
 LTNAEGIMDSPHRLILYVISGAHEIDPRFFFDTLSYQQILQT 119
 TCGTTGATTTAAATTTAACGGGAGTCATGGTAACGGTCAACGGGTTACCAATTGAGTTGGATTATTGATGTTGTCATACGATTAATAACAAATTGAAACGGGGGTTACTGAA 30240
 TS.KFNNLPLRITVQPNNGISTSPNHIQINNMWRNIVISRPTVQI 79

TAGACGGGTTCTGTAGCTTGGTGTACATGATCGGTTGTTGTCAGACCTCCAAAGCAGGGCAATTGTTAAATCCGACTCCAATTCCGAAGAAGCCGGAGATAAGTC 30360
S A T E T R K P P V H I P Q K N L G G F R P G I T L D R S W N G F F G P A Y T M 39

TAGAAGCCGGTCCCTATTTGAAATAAACGGTTATTCCTAAAAGACTGATATTAGTTCACATACGGTTGTCGTTAAAGTAAAAGCAGGGTGTGGTGTGACTCCCCATAGCTG 30460
H L G N G I Q I F R N N G L L S I N T G C L T Q E N L T F H S N T P Q S G M 1

AGGGGTTAAATTACACACAATGCACTCGTACCGTGGACTATCTGAAATGTTGCCGGGTATGTGACACATTACAGTCGTTACCGTTATATAATGTTAGGTAGGAGGCCAT 30600
AAAAATTTGATGGCGTAAAGGTTCTCACTTACCGTGACCTTTTATTACATCGGTTTATTGATGTTACATTGTCCTTICATCCGGACGGATGACCTTTTC 30720

21 N E E P I C Y D T O K L L D D L S N L K Y Q E A D N E 27
ATATCACGTTAAAGTCACTGGTAGAATACCATGGAGAACCAATTGTTATGATACACAAAACCTTGGATGTTAAAGTAATGAAAGTACAAGAAGCAGGACAACGAAA 30840
R P W S P E K T E I A R V K V V K F L R S T O K I P A K H F I Q I N E P L H S N 67
GACCATGGTACCCAGAGAAAACAGAAATCGCCAGGTTAAGGTTACGATCTACCCAGAAAATCCAGCTAACATTTTACAGATATGGAAACCCCTGCATTCTAATA 30960
I C F V Y S N T F L A E A A F T A E N L P G L L F W R L D L O W T I E E P G N S 107
TCTGTTGATATCCAAATACATTTGGCGGGCTCTTCACGGCGAAAATTACCCGACTGTTGGGAGACTAGATCTAGACTGGACGATAGAGGAGCCAGGTAATAGCT 31080
L K I L T Q L S S V V Q D S E T L H R L S A N K L R T S S K F G P V S I H F I I 147
AAAAATTTAACCCAGCTATCAAGTGTAGTACAAGATCCGAGCTTACATCGTTATCGGCAAATAATTACGAACCTCTGCTAAATTGGACCGGTTTCGATACCTTCAATAA 31200
T O M I N N M Y E V A L K D A T T A I E S P F T H A R I G W L E S A I A A L T Q H 187
CGGACTGGATAATATGACGGGCGCTTAAAGGATGCAACACAGCCATTGAACTCACCATTCACTACGGCTGATTGGAATGTTGGAAACCCCATGAGCTTAAACACACATA 31320
K F A I I Y D M P F V Q E G I R V L T Q Y A G W L L P F N V M M N O I Q N S S L 227
AATTGGCGATCATTTACGATATGCCATTGTTCAAGAGGGGATCGTGTAAAACAAATATGCAAGGATGGCTTCGGTTTAACTGTTATGTTGGAAATGAGCTTAAACACACATA 31440
T P L T R A L F I I C M I D E Y L T E T P V H S I S E L F A D T V N L I K D E A 267
CTCTCTAACACGAGCCCTTTTATAATCTGATGATTGATATCTCACCGAAACGGCAGTACATAGCATACAGAAATTCTGAGATAGCTAAATTAAATTAATTAAGATGAGGGCT 31560
F V S I E E A V T N P R T V H E S R I S S A L A Y R D P Y V F E T S P G H L A R 307
TCGTATCCATCGAAGAACCGGTAACGAATCCACGAACGGGCAACGGACTACAGGAATTCTCAGCTCTGGCTTATCGAGACCCCTATGTTTGTGAGACATCCCGGGATGCTGCTAGGA 31680
R L R L D N G I N E S N L L S L S T P G I H I E A L H L L N S D P E A E T T S 347
GACTTAGAGACATGGTATATGGGAAAGCACCTCTTATCGTGTCCACCCCGGAATTCAATTGAGGGCTGTTACATTACTAAACTCCGACCCGGAAACGCGAACACATCTG 31800
G S N V A E H T R G I W E K V Q A S T S P S M L I S T L A E S G F T R F S C K L 387
GAAGTAATGAGCAACACACCCCTGGATTGGGGAAAGGTCAGGCTAGTACATCGCTAGTATGTTAAATAGCACCCCTGGGAATCCGGGTTAACAGATTTCATGCAAATTG 31920
L R R F I A H H T L A G F I H G S V V A D E H I T D F Q O T L G C L A L V G G L 427
TACGTGGTTTATTCGTCACCAACACTGGCCGTTTATTCAGGAGCGCTGTTAGCAGACGAGCATATTACAGATTCTAACACAAACACTAGGATGTCGCTTATGTTGGGACTGG 32040
A Y Q L V E T Y A P T T E Y V L T Y T R T V N E T E K R Y E T L L P A L G L P P 467
CATACCAATTAGTGGAAACGTACGGCTCTACTACCGAGTATGTTAAACATACCGGAGCTAAACCGAGACGGAAAACGGTATGAAACGCTATTACCCGCTTACGGATTACACCGG 32160
G G L G Q I M R R C F A P R P L I E S I Q A T R Y I L L N E I S H A E A R E T T 507
GAGCCCTGGGACAATTATCGGGGCTGTTTGTCTCACGACCCCTTATGAAAGTATACAGGACACGGTAATACTACTAAATGAAATTTCACATGAGACGACTAGAGGAGACACAT 32280
Y F K Q T H N Q S S G A L L P Q A G Q S A V R E A V L T W F D L R M D S R W G I 547
ATTTAACACACATAATCAATCTCAGGAGCTTATACACACAGGAACTGGCTACCCGAAACGGCTACTACCTGTTGACCTAGCTAGGATCAAGATGGGTTATA 32400
T P P V D V G H T P P I C Y D P P A T G L E A V M I T E A L K I A Y P T E Y N R 587
CTCCCCGGTGGATGGGTATGACACCTCTTATGTTGTTGATCCACGGCTACAGGGTTGGAGCTGATGATAACAGAAGCCTAAAGATGCTATCTACCGGAATATAACGCT 32520
S S V F V E P S F V P Y I I A T S T L D A L S A T I L S F D T R G I Q Q A L S 627
CTACGGCTGTTGGAAACCGTCGTTGTCCTTATATTGCAACAGCAGCTGATGCCCTTCGGCAACATAGCTTGTCTTGTACACGGGGAAATACAGCAACGGCTGCTA 32640
I L O W A R D Y G S G T V P N A D G Y R T K L S A L I T I L E P F T R T H P P V 667
TTCTTCAGTGGCTCGCCTATTGGATCGGAACCGCTGCCAACATCGAGATGGATATCCACAAAACACTCTGCTCTTAAACATATTAGAACCTTACCGTACACCCCCAGTAC 32760
L L P S H Y S T I D S L I C E L H R T V G I A Y D L L P Q H V R P L V P D R P S 707
TTTACCATCTCAGTTCTACTATAGATTCTCTTATGCAACTCTACGACTGTTGGCTGACCTGCTGCTCTTGGCTGACCTGCTCTTGTGCTCTGACCTGCTCTTCTA 32880
I T N S V F L A T L Y Y D E L Y G R W T R L D K T S Q A L V E N F T S H A L V Y 747
TTACAAATAGCTTGTGCACTCTTATGATGAACTTACGGCTGGACCCACTGGATAAAACATCGCAGGGTGGTTGAAAATTACATCCAACGGCTAGTGGTT 33000
S R Y W L M L Q K F F A C R F Y P T P D L Q A V G I C N K P V E R D E Q F G V W 787
CTCGGGTACATGTTAGTCAACAAATTGGCTGGCTGTTTATCCACCCAGATCTCAGGGCTGTTGTTCTGTAACCCAAAGGTTGACCCGATGAAACATTGGGGTAGCC 33120
R L N D L A D A V G H I V G T I Q G I R T Q N R V G I S S L R T I N A D A S S A 827
GTTAAACGATCTGCTGATGCGTGGTGTGATATTGTTGGACAATACAGGAATCCGAACGCAATGAGAGTGGATAATCCACGGCTGCGCACAATATGGCGATGCTCTCAGCCC 33240
L R E C E N L M T K T S T S A I G P L F S T H A S R Y A R F T O D Q N D I L M R 867
TTAGGGAAATGAAAATTAACTAACACCTCACTCTCTATGGGCCCTTCTTCAACAGGATGCTCCCGTATGCGCTGGTTACACAGGATCAATGGACATTAACTGCGTG 33360

V D K L T T G E N I P G L A N V E I F L N R W E R I A T A C R H A T A V P S A E 907
 TTGACAACTAACACAGGGAAAATATACCGGTCTTGCATAATGTAGAGATTTTAAATAGGGGAACGAATGCAACAGCTGTAGGCATGCCAGGGCAGTCCCGTCGGCGAAT 33480

S I A T V C H E M L R R G L K N I Q E D R V N A P T S Y M S H A R N L E D H K A A 947
 CTATTGCAACCGTGTAATGTAGGGCGGGTTAAAAAAATACAAGGGATCTGTAAATGCCCAACCTCATATATGAGTCACGCCGAATCTGGAAAGATCACAGGCAGCAG 33600

Y S F V W M D S R Q Q F I V O S G P Q M G A V L T S Q C N I G T W E H Y N A T F L 987
 TTTCATTCGTTATGGACTCCAGCACAGTTATGTGATTCTGACCTCAGATGGCCGGGTTTAACTTCACATGTAAATAGGAACATGGGAATGTAAATGCAACGTTTAC 33720

H D N V K I T T T V R D V I S E A P T L I I G Q R N L R P D E I L S N V D L R L 1027
 ATGATAATGTAAATAACGACAACGGCTAGAGACGTAATTTCAGAGGCTCCGACGCTGATAATAGGACAAGATGGCTCGCCAGATGAGATTTATCTAATGTAGATTTCGCTTG 33840

G V P G N T S G S D P - * 1038
 CGGTACCCGGAAACAGTGGAGTGACCTTAATATAAAACAGGGCTGTTTATGTACATTAAGTATTGTGTTTATTGTAGGGCTTCGTTGATAACCGCTGGCTG 33960

GTATTTCTACACCTCTAGGTTTGGAGCTACACGTCTTGGGATTGAAATCGTAACGCTACGGCTCCCTACAGTGTAGCCGTAATTTCAGCAATA 34080

22 M D I I P P I A V T V A G V G S R H Q F D G A L G P A S G L S C L R T S L S F 39
 AAATGGATAATTCGGCTATAGCTGCACTGTGCGGAGTGGGAAGCCGTAATCAATTGCGCTGGGACCGGCTCAGGTCTGTCTAGTTAAAGACATCTTATCGTTT 34200

L H M T Y A H G I N A T L S S D M I D G C L O E G A A N T T D L S N M G R G Y P 79
 TGCAATGACATATGGCATGAAATGCAACCCCTGTCAGACATGATGGATGTTACAAGAGGGTGCAGCATGGACTACGGATCTGCTAATATGGGAGGGTGTCCCAG 34320

D M C A L V D L P N R I S Y I K L G D T T S T C C V L S R I Y G D S H F F T V P 119
 ATATGTGCTGCTGTTGATCTCCCAATCGAATTTCATATATAAACTGGGGACACTACAGTAGCTGTCGTTGTTGAGATAATACCGGATAGCCATTTCGCTTACCGT 34440

D E G F M C T Q I P A R A F F F D D V W M G R E E S Y T I I T V D S T G M A I Y R 159
 ACAGGGTTTATGTGACACAAATTCCGCTAGAGCTTCTGATGATGTTGGAGACGCTGAGAGCTGATACTAACTYGTAGACTACCGGAATGGCCTATCGTC 34560

Q G N I S F I F D P H G H G T I G Q A V V V R V N T D V Y S Y I A S E Y T H R 199
 AGGGAAACATATCTTATTTGATCCACATGGCCATGGGACTATAGGACAGGCTGAGTTGCTGGGATACACCGGATGTACTCTTATATCGCATGGAGTACCCACCGC 34680

P D N V E S Q W A A A A L V F F V T A N D G P V S E E A L S S A V T L I Y G S C D 239
 CCGATAACCTAGAACATGGCCGCGTCGATTAGTTTGTACCGCAACACGCGTCCGTAAGCGAAGGGCTCTTCGGCAGTAACGCTTATACCGGAACGTGTGATA 34800

T Y F T D E Q Y C E K L V T A Q H P L L L S P P H N S T T I V L N K S S I V P L H 279
 CATATTTACAGATGAAACATATTGCGAAAACCTGGTCAAGCTCAGCTCAACATCCGCTCTTCACCTCTAATCCACGACAATTGCTTAATAATCGCTATAGTACCTCTTAC 34920

Q N V G E S Y S L E A T L H S T L T N T V A L D P R C S Y S E V D P N H A V L E 319
 AAAACGTTGCTGAAAGTGTATCTCTGAGCAACCCCTACATTCACGTTAACCAACCGGTTGCACTGGACCTAGTGTAGTACACCGGAGTGTACTCTGGCATGGCTAGAAA 35040

T T S T G S G V L D C R R R R P S W T P P S S E E N L A C I D D G L V N N T H 359
 CAACCTGACTGGCTGGCTGGTTGGATGTGCTGTAAGCAGCGCCCTCTCATGGACTCTCCCTCAAGCGAGGAAAATTGCTGTATCGCATGGCTGGTAAATAACACATT 35160

S T D N L H K P A K K Y L K F K P T V D Y P D K T Q V A H Y L P R L R E V A N T 399
 CCACGATAATTCACATACCCGCTAAAGGTTCTCAATTAAACCAACTGTAGACGTCGCGGATAAAACACAGGGACATGTTACCCCGCTACGAGAAGTGTGTAACACCC 35280

P D V V L N Y S N V D T P E S S P T F S R N M N V G S S L K D R K P F L F E Q S 439
 CAGACGTTGTTAAATGTATCCATGTAGATAACGCTGAACTCCAGTCCACTTTCACGGAACATGAAATGTAGGAAAGCAGTTGAAAGATCGGAAGCCATTCTATTGAAACAGAGT 35400

G D V N M V V E K L L Q H G H E I S N G Y V Q N A V G T L O T V I T G H T N V P 479
 GTGATGTCACATGGTGTGCAAAACTACTAACACATGGCATGAAATTGCAATGTTACGTAACAAATGGGGGGTACCTGGGATACTGTATACCGGTACATAAAATGTTCCA 35520

I N V T R P L V M P D E K O P L E L F I N L T I L R L T G F V V E N G T R T H H 519
 TTGGGTAACAGCCCTGGTTATGCCAGACGAAAAGGATCCATTGGAGCTTTTAAACCTCACCTTGGCTTAAACGGGATTGTTGGGAAATGGAACACGTACACATCG 35640

G A T S V V S D F I G P L G E I L T G F P S A A E L I R Y T S L I L T N M P G A 559
 GTGCTACAGCGTTGATCAGACTTATAGGCCCCCTGGGAAATTAAACAGGATTCGCCCTCCGCCGGAACTTACCGCTTAACAGTTGATATAACAAACATGCCGGGGCG 35760

E Y A I K T V L R K K C T I G M L I I A K F G L V A M R V Q D T T G A L H A E L 599
 AAATGCTATTAACCTGGCTCCGGAAAAAATGTACAATTGGCATGCTATTACGCTAAGTTGGTAGTGTGCTGAGGTTACGGGTTACAGGATAACCCGGCTTACATGCCGAACACTG 35880

D V L E A D L G G S S P I D L Y S R L S T G L I S N S P I I S H P G L F A E 639
 ATGTGTTAGAAGGGATCTAGGGGTGTCGCCCCATAGACCTCTATTCTAGACTGTGTCAGGGCTTATAAGTATACTAAATTGCTTATTCATCCGGACTTTTGGCGAGC 36000

L I P T R T G S L S E R I R L L C E L V S A R E T R Y M R E H T A L V S S V K A 679
 TTATTCACCCGCTACGGGCTCCCTGTCGAACGAAATAGCTCTTGTGAAATTAGTCTCGGCCCCGGAGACCGCTATATCGCTGAAACACCGCCGTTGTTCTAGTGTAAAGGCTT 36120

L E N A L R S T R N K I D A I Q I P E V P Q E P P E E T D I P P E E L I R R V Y 719
 TAGAGAATGCTACGGCTACCCGCAATAATTGATGCCATTCAAATACAGGAAGTCCCGGAGACGGAAACCCCGGAGAACCCGACATTCCACCGGAAGAGTTAATCGCGCTGTATATG 36240

E I R S E V T H L L T S A V T E Y F T R G V L Y S T R A L I A E Q S P R R F R V 759
 AGATACGATCGGAAGTACAATGCTATTGACCTCGGCTTACAGAATACTTCACCCGGAGTGTATATAGCACACGGCCCTGATCGCTGAAACATCCCTAGGCCTGGCTG 36360

A T A S T A P I Q R L L D S L P E F O A K L T A I I S S L S I H P P P E T I Q N 799
 CGACCGCAAGTACGGCACCCATTCAACGGCTTTAGATTCCTCCGGAATTGACGCTAAATTAACGCCAATCATATCGTCCTGCTATAACCCCTCCCTGAGACTATAACAAATC 36480
 L P V V S L L K E L I K E G E D L N T D T A L V S W L S V V G E A Q T A G Y L S 839
 TCCCCGTGATCTCTGTTAAAAGAGCTTAAAGAAGGGGAAGATTAAACACAGACACGGCTCTGTATCGGGTATCTGTAGTCGGGGAGGCTCAAACCGCAGGTTACTTATCCA 36600
 R R E F O E L S R T I K T I N T R A T Q R A S A E A E L S C F N T L S A A Y D Q 879
 GACGAGAGTCGATGAAATTACCGTACAATTAAACCTTAATACACGCCAACGCCACGGGCTCCCGCGAAGCAGAGTGTCTGCTTAACGCTAACGGCGCGTAGACCAAG 36720
 A V K D Y E T Y N G E V K Y P E I T R O D L L A T I V R A T D D L V R O I K I 919
 CCCGAAAGGACTATGAAACATATAACATGGTGAGGCTCAAGTATCTGAAATAACACGGGATGTTTATAGCAACAAATTGTACGTCTACAGCAGATTGGCCACAGATAAAATT 36840
 L S D P H I Q S G L O P S I K R R L E T R L K E V Q T Y A N E A R T T O D T I K 959
 TAAGTGATCCAATGATCCAATCCGTTTACAACCTTCGATAAAAGCCATTGGAAACAAGCCTTAAAGAGGTTCAAGCCTGATCCAAACAGGGCCGAAICCACACAGGACACAATAAGA 36960
 S R K Q A A Y H N K L G G L L R P V T G F V G L R A A Y D L L P E L A S E L D V Q 999
 GTGAAACAGGGCGCATATAAAACTGGGGGTTACTTCGCGGTTAACCGGTTTGTGGGACTTAGGGCTCAGTAGATTAAACCGGAACTTCTGAGTTAGATGTCCAAG 37080
 G A L V N L R T K V L E A P V E I R S Q L T G D F W A L F N Q Y R D I L E H P G 1039
 GAGCCCTGGTAAATCTCAGGACCAAAGCTTAGAGGGCGGTAGAGATCCGTTCTCAACTTACGGGTGATTCTGGCGTTATTAACCAATATCGAGACATTTCAGGAACTCCGGAA 37200
 N A R T S V L G G L G A C F T A T I I E I V P I P T E Y R P S L L A F F G D V A D 1079
 AGCGACACATCTGCTTGGAGGACTGGGACTTGTGTTACAGCTTATTCGAAATGTGCGGATCCTACGGGATATAGGACATCATTCGCTGGTTTTGGTACGTGGCAGATG 37320
 V L A S D I A T V S T N P E S E S A I N A V V A T L S K A T L V S S T Y P A L S 1119
 TGCTTGATCCGACATCGCAGCCGATCTACTAACCCGGAAAGTGAAGCTCCGATAACCGCTGTTGCAACTCTAGAAAGCAGCTTAGTTACAGTGCAGCTTACAGTGCAGCTTATCCT 37440
 F V L S L Y K K Y Q A L Q O E I T N T H K L T E L O K Q L G D D F S T L A V S S 1159
 TTGTTGTCGTTATATAAAATATCAGGCTTACACAAGAAATTACAAGATAACCCATAAGTGAATTACAAAACACTGGAGATGTTCTCCACCCCTAGCTGCTCATCTG 37560
 G H L K F I S S S N V D D Y E I N D A I L S I Q T N Y H A L M D T V K L V E V E 1199
 GACACTTGAGTTTATCATCTCAAATGAGATTTAGAAATAACCGATGGGATTTACATAACAAACAAATGTCACCCCTAATGGGATACGGGTTAACCTGTTGAGTTGAAC 37680
 L Q K L P P H C I A G T S T L S R Y V K D L H K L V T M A H E K K E Q A K V L I 1239
 TCCAAAAGCTACCCCCCATTGTTATGCTGGGACATCTACCTTATCTCGAGTAGATAAGGGATCTCATAAACCTGTCACATGGAGAAGGAACGGCAAAGTTAAATT 37800
 T D C E R A H K Q O T T R V L Y E R W T R D I I A C L E A M E T R H I F N G T E 1279
 CGGATTGACGTGACATACAAACAAACGACTCGGGTTTGATATAACCGAGTTACCCACAAGCAGCTGGGTAGCGGTGATTTAGCATGTCCTGGGCAATGGAAACCGCCATATAACCGGACAGAAC 37920
 L A R L R D M A A A G G F D I H A V Y P Q A R Q V V A A C E T T A V T A L D T V 1319
 TGGCACGGTTGGAGATATGGCGCTGGGGAGGGTTGATATAACCGAGTTACCCACAAGCAGCTGGGTAGCGGTGATGAAACTACGCCATTAGGCTTACGGGATAGATACTGTG 38040
 F R H N P Y T P E N T N I P P P L A L L R G L T W F D D F S I T A P Y F T V M F 1359
 TTGGCCACAATCCATATCCCCGAAATAACAAATTCACCCACCTTGGCTTGTGTTAGAGGGGTTAACATGGTTGATGTTTCGATTACCGCTCCGTTACCGTTATGTT 38160
 P G V S I E G L L L L M R I R A V Y L L S A D T S I N G I P N Y R D N I L R T S 1399
 CAGGTGTTAGATTGAGGACTCTCTCTGTTATCGCTGGGTTGTTATACCGGCTTATGAGTTAACACGGCTTATGAGTTAACCTGAGATACTGGGATATGTTACCGAACCTGG 38280
 G D L L Q I P A L A G Y V D F Y T R S Y D Q F I T E S V T L S E L R A D I R Q A 1439
 GGGACTTACAAATACCCGATTTGGCTGGGATTTGATTTACACCGCTTATGAGTTAACACGGCTTATGAGTTAACCTGAGATACTGGGATATGTTACCGAACCTGG 38400
 A G A K L T E A N K A L E E V T H Y R A H E T A K L A L K E G V F I T L P S E G 1479
 CGGGGCTAACTTACAGAACAAATAACGGCTTGGAGGAGTACTCATGTCGGGCAACAGAACCGCTAACTTCGACTAAAGAGGTTGCTTCATTTACACGGAAAGGT 38520
 L L I R A I E Y F T T F D H K R F I G T A Y E R V L Q T M V D R D L K E A N A E 1519
 TATTGATTCGGGCTATAGAGTTTACACTTCGATCATAAACGATTATAGGAACGGCATATAAGAGGTTTACACAAACATGGTAGACGGCTCATTAAGGAGGGCAACCGAGAC 38640
 L A Q F R M V C Q A T K N R A I Q I L Q N I V D T A M A T E Q Q E O Y D F T N L 1559
 TTGACAGTTCTGTTGCTGAGGCAACAAAGAACGGCTAACAAATTTACAAAATGTTGATACGGCAATGCCACTGAGCAACAGAACGTTGATTCTACTAACCTGA 38760
 K T L L K L T P P P P K T I A L A I D R S T S V Q D I V T O F A L L L G R L E E E 1599
 AGACGTTATTAACCCCTCCAAAACATTGCGATGAGCTACTTCGCTCAGGACATTGTCAGGCTTGTGCTAGCTGAGCTTGTGCTAGCTGCTGTTAGGGCTCTGGAAAGAAA 38880
 T G T L D I Q A V D W M Y Q A R N I I D S H P L S V R I D G T G P L H T Y K D R 1639
 CTGGTACGTTGGACATTAGCTGGCTGTTGACTGGATGACCAAGCTGGCAATTATGACTCCCATCAACTAGTGTGCTATAGCGGATACGGCCCTGCTACATTAACAGATGG 39000
 V D K L Y A L R T K L D L L R R R I E T G E V T W D D A M T T F K R E T G D M L 1679
 TGGATAAACTTATGCGTACGCAACTAAATTGAGCTCTACAGGAGAATAGAACCGGTTAGCTGGGAGGTTACGTTGAGCTGAGCAACATTAAAAGAGAAACGGGGATATGTTG 39120
 A S G D T Y A T S V D S I K A L Q A S A S V V D M L C S E P E F F L L P V E T K 1719
 CATCGGGGACGCTACGCTACTTCGCTAGATGATAACAGGACTCCAGGATCGGCTCTGTTGACATGCTTGTGCTAGCTGCTGTTAGGGCTCTGGAAAGAAA 39240
 N R L O K K O Q E R K T A L D V V L Q K Q R Q F E E T A S R L R A L I E R I P T 1759
 ACCGCTCCAAAAAGAACAGGAAACGTAACCGCGTTGGATGTTGTTGCAAAACAAAGACAGTTGAAGAGACCCGCTCTGCTTACGAGCTTAAATTGAGCTTACCGAACCG 39360

E S D H D V L R M L L R D F D O F T H L P I N I K T Q Y M T F R N L L M V R L G 1799
 AGAGTGCACCATGAGCTTCCTGATGTTTACGTGATTCGATCAATTACATTCGCTATGGATAAAAACACAGTATAGCATTTCGAAATTACTCGATGGTACGGTTAGGCT 39480
 L Y A S Y A E I F P P A S P N G V F A P I P A M S G V C L E D Q S R C I R A R V 1839
 TGATGCAAGTTATGCTGAGATTTCACCCGGCTCCAAACGGAGTTTGCTCTATTCGGGTTATGCTAGAGACCAATCCGATGCATTCCGGCGGGTGG 39600
 A A F M G E A S V V Q T F R E A R S S I D A L F G K N L T F Y L O T D G V P L R 1879
 CGCGGTTATGGGGAGGGCTCTGTTGCAAACGTTAGGGAAAGCCAGATCTCTATAGACGCTTGTGTTGGAAAAAATTTAACCTTTACTGGATACTGATGGGTTCCACTTCGAT 39720
 Y R V C Y K S V G V K L G T M L C S Q G G L S L R P A L P D E G I V E E T T L S 1919
 ATAGAGTGTGTTAAAATCAGTTGGGGTAAACTTGGGAACCATGCTATGCACTGGGTTATCTTACGACCGGACTTCGGATAGGTTATGTGGAAAGAAACTACACTTCG 39840
 A L R V A N E V H E L R I E Y E S A I K S G F S A F S T F V R H R H A E W G K T 1959
 CATTACGGCTGCCAATGAGGCTAACGATGAGCTACGCAATTAGAATCCGGTTTCTGCTTCCACCTTGTGATGGCATTGGGAAACCCGGATGGGTAACCA 39960
 N A R R A I A E I V A G L I T T T L T R O Y G V H W D K L I Y S F E K H H L T S 1999
 ACACCCAGCAGGCATTGCAAGAGATAACCCGGCCCTATAACAACACATTGACACGACAATACGGGTTCTGGGACAAGCTTATTCTTGTGAAAACACCCACCTAACCTCTG 40080
 V M G N G L T K P I Q R R G D V R V L E L T L S D I V T I L V A T T P V H L L N 2039
 TAATGGGAACTGGACTAACTAACCAATCAGAGAAGGGGTGATGTCAGCGTATTAGCTTACCTGATATTGTAACTTTGGTTGCAACACCCGGTACATCTCAATT 40200
 F A R L D L I K Q H E Y M A R T L R P V I E A A F R G R L L V R S L D G D P K G 2079
 TTGCTAGATGGGTTAATTAACAGCATGAGTATGGCCGCTACCCCTGACAGCGTAACTGGGCGATTAGGGTCGTTACTCGTCGCTATTGGATGGAGACCCGAAAGGCA 40320
 N A R A F F N A A P S K H K L P L A L G S N Q D P T G G R I F A F R M A D W K L 2119
 ATGCCGGGGCTTTTAAATGCCGCCCTACCAACATAAACCTCCGTTAGCTGATCAACAAAGATCTACCCGGGGGAAATTGGCATTGGCATGGCAGATTGGAAACTTG 40440
 V K N P Q K I T D P F A P M O L S P P P G V K A N Y D A V T R I M A T D R L A T 2159
 TYAAATGCCACAGAAAATAACGGATCCTTTCGCGCATGGCAATTCCCCCCCCCCCCGGGTAAGGCCAATGTCGATGGCAGTTACCGGTATAATGGCACAGATCGCTTGGCGACCA 40560
 I T V L G R M C L P P I S L V S N W N T L Q P E E F A Y R T Q D D Y D I I V D A 2199
 TTACTGACTTGGGCGCATGTCCTCCGCCATTCTCTTAGTGTCAATGTCGAAACCGCTGCAACAGGAGATTGGCATTGGCATACAGACACAAAGATGATGTGACATTATGATGGCA 40680
 R L D L S S T L N A R F D T A P S N T T L E W N T D R K V I T D A Y I Q T G A T 2239
 GACTGGATTGTCATCCACGCTTAATGCAAGATTGATACCGCTCCAGCAATACCGCTTAGAGTGGAAATACAGCGTAAAGTATTACAGATGCTTATTCAAACCGGGCAACGA 40800
 T V F T V T G A A P T H V S N V T A F D I A T T A I L F G A P L V I A M E L T S 2279
 CAGTTTACAGTAACGGGGCGGACCAACTCACGTTCTAATGTAACAGCGTTGACATGCAACTACGGCTATTGGGCTCCCTGGTTATGGCATGGAACTTACATCG 40920
 V F S Q N S G L T L G L K L F D S R H W A T D S G I S S A V S P D I V S W G L R 2319
 TTTTTCACAAATTCCGACTTACTTGGGTTAAATTATCCTGATCCGGCATATGGCTACAGATTCGGGTATTCCTCAGCGTATCTCCGATATTGGCTTGGGGGTACGTT 41040
 L L H M D P H P I E N A C L I V Q L E K L S A L I A N K P L T H N P P C L L L L 2359
 TACTGCAATGGCATCTCACCAATTGAAATGCACTGTTAATGTCCTAACAGAAAATCTGCGCGCTATTGCAAAACCTCTACAAACATCCCCGTGTTACTGCTATGG 41160
 D E H M N P S Y V L W E R K D S I P A P D Y V V F W G P E S L I D L P Y I D S D 2399
 ACGAACATATGATACCTCTTGTGTTATGGGAAACGAAAAGACTGATTCCAGCTCCGGATTATGGGCTTTGGGGGCAAGATCTTATGATTTGGCTACATCGACTCCGATG 41280
 E D S F P S C P D D P F Y S Q I I A G Y A P Q G P P N L D T T D F Y P T E P L F 2439
 AGGACTCTTCCCTCGTGTCCGATGATCCATTACTCGCAATTATGGCGGTATGCGCCCAAGGCCCCAAACCTGACACAACTGATTTTACCAACGGGACCACTATTA 41400
 K S P V Q V V R S S K C K K W P V R P A Q P A Q P A Q P A Q T V Q P A Q P 2479
 AGTCTCCCGTCAAGTTGTTAGAGTCAATGAAAAAAATGCCGCTCCGGCAGCCCCGGCAGCCCCGGCAGCCCCGGCAGACGGTCCAGCCCCGGCAGCCCC 41520
 < A X A X A X A X A X B X A X B >
 reiteration R3
 I E P G T Q I V V Q N F K K P Q S V K T T L S Q K D I P L Y V E T E S E T A V L 2519
 TAGAACCGGGCACAAATAGTGGYACAAACCCAAACCGTAAACCAACCTTGTGCAAAAGATTCCTCTGATGGAAACGGAACTCAGAACCGCTGTGCTTA 41640
 I P K Q L T T S I K T T V C K S I T T P P N N Q L S D M K N N P Q N Q T L N Q A 2559
 TACCTAAGCAATTAACCCCTCCATTAAACACCGTTGTAAGTATTACCCACCAAATAACCAATTGTCGCGATTGGAAAAATAATCCACAGCAAACCAACGTTAAACCAACGCT 41760
 F S K P I L E I T S I P T D D S I S Y R T W I E K S N Q T Q K R H O N D P R M Y 2599
 TCAGTAAACCAACTTGAGGATACCTGCAAGACTGACTGCAATTCTACCGACTGGATTGAAATCAATGAAACACAAAAACGGCATCAAATGACCCCTGAAATGTATA 41880
 N S K T V F H P V N N Q L P S M V D T A A D A P Q T D O L L T H Y K T R Q P S P H 2639
 ACTCCAAAACAGTATTCCACCCCTGTAATAACCAATTACCTCTGGGTGACACGGCAGCCGATGCCCTAACGGACCTATTGACAACAACTATAAAACAGACAGCGTCGCCAAACT 42000
 F P R D V H T W G V S S N P F N S P N R D L Y Q S S D F S E P S O D G Y S S E E N 2679
 TTCCCGGGGAGCTACACCATGGGGCTATCTCTAACCCGTTAACTCACCGAACAGAACGACCTATCAAAAGTGTGTTTACTGACCGCTATGCGATGGAGTGAATT 42120
 S I V L S L D E H R S C R V P R H V R V V N A D V V T G R R Y V R G T A L G A L 2719
 CTATCGTACTAAGTCGACGACATCGGCTATGTCGCGTCTCTAGGCACGTCAGCGCTGTTATGCGATGAGTGTGACCGCTTGTGCGATGGGGACCCCTGGGACTGG 42240
 A L L S Q A C R R H I D N V R Y T R K L L H D H T E D I F Q G L G Y V K L L L D 2759
 CACTGTTAACCGGGCATGTCGGCTATGATGACACGTTAGATACAGTAAACCTTAAATGGACCACCGAAGATATTCAGGCTGGGTATGTTAAATTGTTAGATG 42360

2763 R K S F Y F N Q D G L T G G P Q P L A A A L A N M K D C A R M V D C S S S E H R 285
 12480 CGTAAGTCATTTACTTTAATCAGGACGGACTGACTGGAGGCCCTCAACCTTAACTGGCGCCGCTTGGCTAATTGGAAGATGGCGCGGATGGTACTGTCATCATCGAACATCGC 45360
 220
 12600 T S G H I T C A E R A L K E D I E F E D I L I D K L K K S S Y V E A A W G Y A D 325
 180 A C A A G T G G G A T G A T T A C C T G C C C G G A C C G T C A T T A A A A G G G A T A T A G A G G T T A G A G G A T A T A T A G A C A A C T T A A A A A C T G C T T A C G T A G A A G C A G C T T G G G G T T A C G C A G A C 45480
 12720 L A L L L L S G Y V A T W H N V D E R T H C A I E T R Y G C V K S Y W Q A N R I E N 365
 140 T T G C T T I T T A T T A C T G A T G T G G G T T G C T A C T T G G H A T G T A G A C G A C G G T C A A A T T G T G C T A T A G A A A C T G C G G T T G G A T G T G T T A A T C A T A C T G C C A G G C G A C C C G A T T G A A A C 45600
 12840 S R D Y P K Q F S K F T S E D A C P E V A F G P I L L T T L K N A K C R G R T N 405
 100 T C C A G G G A C T T C C A A A C A T T T C C A A T T A C G A C G G C G G A G G A T G C C T G C C G C A G T T T G G C T A T T T G T T A C T A C C T T A A A A A C G C A A A G T G C C T G G C C A C G A A T 45720
 2960 T E C H L C C L L T I G H Y W I A L R Q F K R D I L A Y S A H N T S L F D C I E 445
 60 A C C G A A T G C A T G T T A T G T G T T T A T T A C C A T A G G G C A C T A T T G G A T C C T G C C G C A G T T T A A A A G G G A T A T A T A G C A T A C T C A G C A A A T A C A C A A G T T A T T G A C T G T A T C G A A 45840
 3080 P Y I N A M S L D N P I K L K F P F N D E G R F I T I V K A A G S E A V Y K H L 485
 20 C C T G T A A T C A T G C A T G G A C C C T A G A T A A C C C C A T T A A C C C T A T T C C A T T A T A G T A G G G G T C G A T C A A C C A T T G T A A A A G C A G C A G G T T C C G A G G C C G T A T A A A C A T T A 45960
 3200 F C D O L L C A L S E L Q T N P K I L F A H P T T A D K E V L E L Y K A Q L A A Q 525
 1 T T T G C G A T C T C C T A T G C C T C T C G G A T T A C A G A C A A C C C T A A A T T T G G C C A T C C T A A C C C G G G A T A A G G A A G T G T G G A T T T A A A G C C C A A C T G G C T G C A C A A 46080
 3320 27 M H L K P T R F F H A N Q P P M P H S Y E N E D L 25
 235 N R F E G R Y C A G L W T L A Y A F K A Y Q I F P R K P T A N A A F I R D G G L 565
 A C A G A T T G A A G G T C G T G T A T G T G C T G G C T A T G C A T T A A G C C T A C C A G A T T T C C A C G C A A C C A A C C G C A A T G C C G C A T T C A T A C G A G A T G G A G G A C T T 46200
 3440 C F D D M Q Y R W S P S N T P Y R S M S R R Y K S V S R S G P S M R V R S R T P 65
 195 M L R R H A I S L V S L E H T L S K Y V - 585
 A T G C T T C G A C G A C A T A T C G C T G G T C T C C C T C G A C A C C C T A T C G A A G T A T G T C T A G G C G A T A T A A T C G T A T C T C G G A C G G C C T T C G A T G C G T G T A C G C T C C A G A A C G C C 46320
 3560 155 C R R Q T I R G K L M S K E R S V Y R H Y F N Y I A R S P P E L A T V R G L I 105
 A T G C C C C G T C A A C C A T T C G G A G G G A A A C T T A T G T C A A A G G A C G G T C T G T A C C G C A T T T T A A T C A T G C A A G G T T C C C C C C C A G A A G A C T A T G C T A C C G T T A A T G A G G C T T A A T 46440
 1680 115 V P I I K T T P Y T L P F N L G Q T V A D N C L S L S G M G Y H L G L G G Y C P 145
 C G T G C C A A T T A A T G A C G A C C C C T G C A C C C T C C G T T A C T T G G G C A G A C T G C C T G C G T T A T C G G A T T G G G T T A T T G G C G G T T A T T G T C C 46560
 1800 T C T A S G E P R L C R T D R A A L I L A Y Y V Q Q L N N I Y E Y R V F L A S I L 185
 75 G A C A T G C A C T C T G G A G A C C C G C T C T A T G C G A C C G A T C T G C G G C T C T G A T A C T A G C A G C C T T A A C A C A T A C A T A C G A A T T C G Y G T G T T C T T G C A T C C A T T T 46680
 1920 35 A L S D R A N N Q A A S A E P L L S V L A Q P E L F F F M Y H I M R E G G W R D 225
 040 G C C G C T A C A G C C G A G C C A C A T G C A A G C C G C T C C C T A T T G C G A C C G G A A T T T T T T A T G T A T C A T A T A T G A G G G A G G G G G C A T G C G A G A 46800
 1 A I R V L F Y R D G D A G G F W M M Y V I F P G K S V H L H Y R L I D H I Q A A C R 265
 160 T A T C C G C T A C T T T T A T C G T G A T G G A G A T G C C G G A G G G T T A T G A T G T A T G T A T T T C G G G G A A A C T G T C A C C T C C A T T C A G A C T A T C G A T C A T A C A G G C C G T G T C G 46920
 154 G Y K I V A H V W Q T T F L L S V C R N P E Q Q T E T V V P S I G T S D V Y C K 305
 G G G T T A A A A T A G T C G C A C A C G T T G G C A G A C A C A T T T T A C T G T C G G T A T G T C G C A A C C C A G A C A C A C A G A G A C T G T G G T G C C A T C C A T T G G A A C A T C G G A C G T T A C T G T A A 47040
 280 114 N C D L N F D G E L L L E Y K R L Y A L F D D F V P P R - 333
 400 A A T G T G A C C T T A C T T T G A T G G G A A T A C A A A G A C T C A C G C T A T T T G A T G A T T G T C C T C C T C G G T A T T G C G T C A G T G T C A T T T T A T T C C C A G 47160
 74 28 - S Q H L I A K P I C F V R R M I Q H S O E E E T I E A E T H N M K H D W 1160
 *
 5 C A C G G G C G T G T A T A C A A A C A A G C C T G C C G C T G C A A G C G G T T A C G T T A C A C T C G T G T C T G G A A T A A A A C G T T T A A A G C C G T T C T G T G A T G T T A G T G T C G T T 47280
 520 C P A H I C V F G A A Q L R N L M K V N V V R T E P I F R K L L R E T L K T D N 1120
 34 C C A A T A A C C C T T A A A G T T A C A T C G C G C T C C A T G A G A T G A G A A A A T A A T A G T C A T A G T T T A A A G C A G C C C G T G T G A T G T C A G T G G A T C T C C G C T A A G T C A G A T T 47400
 45 G F L A K F T V S A T G I L H S F Y Y D I N L S L G H S T V H I P D E A L O S I 1080
 340 A T T A A C T T A C G C T T C C C T C C C A C C G C T T A C T C G C G T A T T C T G T G A A G G A T C T C C A C G T A G C A A G C T A C A T T T G C A T C G C C T C A C T C G C T G T G G G G C C A A A C A 47520
 1 I L K R K A E G C R K G A T N Q L P D G R L L A V S K A D A E V E D T P A V I Y 1040
 85 T A A G G G A T G C G T T C C G A A C C T T G G G A T T G A C C C T G T C T C A T T A T A T A T A C T G T G A A G G A T C T C C A C G T A G C A A G C T A C A T T T G C A T C G C C T C A C T C G C T G T G G G G C C A A A C A 47640
 60 Y P I R E R V N P I Q G Q R M V L K Y Y Y V T L H A L R R N I Y A S P P R S L E A 1000
 125 G T C A T A C C A A C T T A T T A T C C A T T T G G G T G A T G T A A T C T G G C G A T G T C G C A T T A T G C G T C C A A C C C G G C C A T C C C A G C G G C A T G G C C C G T C T A T T C C A T T C A G C A A T G 47760
 80 T H V F K N I D L K P S T I Q R H A D A I I R G F G A M G S P N A R R N M E A I 960
 165 G A A A C A C A C G C G C T C C G C G C A G C A C G C G A G A C G G T G C G T C A T A A C A C A G T T C T C A G A C T G T G G G C A T A T C C T A A T T C A G A C T G T G T T T T C T A C C A A G T C G A C T 47880
 100 S V C S A E A A A R S V T D D Y L L L E V L K R A Y D N I F Q C N N K R V L D V 920
 120 C C C T C A T T A A C C T T C C C G C G T A A A T C C C C A A T G T A C T T T G T T A A G C A A A G T T T A T A A A A G T T T T C A C A C T C C A A C T T T A T A G G G A G C A A A C A G G C C G T T 48000
 140 G K M L V K G Y I V G I Y K K K T I L L L K I F T K E C E L K I P P C F L A T 880
 145 G A A T T A T A G T G C C A T T T C G C G G A T T T A G C T A T C C C C T C A A C A C T A A C C C C T G A T C C G G A T A A C A C A G A T C C G T A T C T C C A T A T A A C C T T A C C T C G T A C C G T T T T G G 48120
 140 S I I H A N K E G I K A I G E V S V G K F R I F V S D T O G Y I V K V E Y A K Q 840

P S V F V F H G G K H V L P S S A A P N L T R A C H A A R E R F G F S R C Q G P 128
C C A T C G T C T T T C G C G G G C A A C G C T T T C A C C G C G G C C C C A A T C A C C G C G G C T G T A C C G C G G C T C G A G A C G G T T G G G T T C A C G C T G C C A A G G C C T 51240

P V D G A V E T T G A E I C T R L G L E P E N T I L Y L V V T A L F K E A V F M 168
C C T G T G A C G G T G C T G T G A C G A C G C G C C G T G A G A T A G C A C C G C C T T G G A T T A G G C C A G A A A A T C A C A T T A T T A C T T G G G T G C A C G C T A G T G T T A A C C G G C C T G A C C G C T 51360

C H Y F L H Y G G L D I V H I N H G D V I R I P L F P V Q L F M P D V N R L V P 208
T G C A A C G T G T T C T G C A T T A T G G A G G A C T C G A T T G T T C A T A C C A T G G G G A T G T T A C C G T A T A C C G T T A T T C C G G T A C A C T T I C A T G C C G A T G T Y A C C G T G C C A C C G C T 51480

D P F N T H H R S I G E G F V V P T P F Y N T G L C H I H D C V I A P M A V A 248
G A C C C A T T C A C A C T C A C A G G C T C A T C G G A G G G T T T G T A Y A C C C A A C C C C T T T A T A C C C G G T T G C C A T T T A T A C T G A C T G T G T T A T T G C T C C A T G C C G C T G C C 51600

L R V R N V T A V A R G A A H L A F D E N H E G A V L P P D I T Y T Y F Q S S S 288
T G C C G C T G A G A A T G T A C T G C C G T G C C C G A G G A G C G C C A C C T T G C T T T G A T G A A A A T C A C G A G G G G G C A T G A C T C C C C C C T G A C T A C G T A C A C G T A T T T C A G T C C T C C T I C A 51720

S G T T T A R G A R R N O V N S T S K P S P S G G G F E R R L A S I M A A D T A L 328
A G T G G A C C A T A C C G C C C G T G G A G C G G T C G A A C G T G T C A C T C C A C G T C T A G G C T A C G G C C A T G G G G G G T T G A A G A C G G T T G G C T C T A T T A G G C E G T G A C A C G C C T G 51840

H A E V I F N T G I Y E E T P T D I K E W P M F I G H E G T L P R L N A L G S Y 368
C A C C G A G A G T T A T T C A C A C T G G A A T T A C G A G A A A C T C C A C A G A T A T C A A G A T G G C C A T G T T A T A G G C A T G G A G G G C A T T T G C C A A G G C T A A C G C T C T G G G G T C A T A T 51960

T A R V A G V I G A M V F S P N S A L Y L T E V E D S G M T E A K D G G P G P S 408
A C C C G C T G T G G C C G G G G T C A T T G G T C C A T G T T T C A G C C C A A T T C T G C T T G T A C T C A C T G A G G T G G A G G A T A G C C G G A T G A C C G A A G G C A T G G G G A C C G G G T C C A T C A 52080

F H R F Y Q F A G P H L A A N P O T D R D G H V L L S S Q S T G S S N T E F S V D 448
T T T A T E G A T T I C A C G T T G C C G G A C C T C A T T T A G G T C G C G A A T C C C C A A C A G A T C G C G A T G G C C A C G T C T A C T C A G G G T T C A T C A A C A C A G G T T A T G C C G G G T C A T A C A A C A G G T T A T G C C G G G T 52200

Y L A L I C G F G A P L L A R L L F Y L E R C D A G A F T T G G H G D A L K Y V T 488
T A T T G G C A C T C A T T G G G A T T T G G A C C A C C C C T G T G G C G G A C T G C T T T T T A T C T G A A G C G T G C G A C T G G C T G C G T T A C A G G G G T C A C G G G G A T C G C T T A A A T A T G I T A C G 52320

G T F D S E I P C S L C E K H T R P V C A H T T V H R L R Q R M P R F G Q A T R 528
G G G A C C T T G A C T C T G A A A T T C C A T G T G A A A A C A C C C G C C G G T A T G C C G T C A C A C A C G A T C A C C G A C T T A G A C A A C G C A T G C C G G A T T T G G A C A A G C C A C C C G T 52440

Q P I G V F G T H N S Q Y S D C D P L G N Y A P Y L I L R K P G D O T E A A K A 568
C A A C C T A T T G G G G T T G G A C A T A G A C A C G C A T A T A G C G A T C G C G A T E C T C A T G A G G A A C T A T G C T C C A T T T A T C T C T C G A A A C C C G G G G A T C A A C G G A A G C G A A A G G C A 52560

T H Q D T Y R A T L E R L F I D L E Q E R L L D R G A P C S S E G L S S V I V D 608
A C C A T G C A G G A C A T T A T G G G C T A C A T G A A G C C T G T T T A T G C A T G A A C A G G C G A C T T A G G A T C C G G G T C C C C A T G T T C T C C G A G G G A T C A T C T C T G T A T T G G G A T 52680

H P T F R R B I L D T L R A R I E Q T T T Q F N K V L V E T R D Y K I R E G L S E 648
C A T C C A C G T T C T G C C A T A T T A G G C T A C A T G A A C A G C A C A C A C A A T T A T G A A G T G T G G T G A G A C C C G G G A T T A A G T C C G T G A A G G T T A T C C G G A A 52800

A T H S M A L T F D P Y S G A F C P C I T N F L V K R T H L A V V Q D L A L S Q C 688
G C C A C C A T T C A T G G C C G T T A C G T T G A T C C A T A C T C A G G A C G T T T G C C C A T T A C C A T T T T A G T T A A C G A A C A C C T A C C C G G G T G A C A A G C T T A G C A T T A G C C A T G T 52920

H C V F Y G Q Q Q V E G R N F R N Q F Q P V L R R R F Y D L F N G G F I S T R S I 728
C A T T G G T T A T T A C G C G A C G A C G T T G G G G G G G A C T T T C G T A C C A A C T C T G T T G C G G G G C G T T T G T G A C C T G T T A T G G G G G G G T T A T C A C A C G C T C T A T A 53040

T V T L S E G P V S A P N P T L G Q D A P A G R T F D G D L A R V S V E V I R D 768
A C C G T A A C T T A T G A A G T C C T G T A T C C G C C C A A T C C G C A T T G G G A C A A G C G C C C G C G G G G C T A C T T G A T G G G G A T T T A G C C G C G T A A G C G T G G A G T T A T C G G G G A T 53160

I R V K N R V V F S G N C T N L S E A A R A R L V G L A S A Y Q R Q E K R V D M 808
A T A C C G A T T A A A T A G G G T C G T T T T C A G G T A C T G A C A A T C T C T G A G G C A G C C C G G C A A G G C T T G A G G C C T T G C A G T C S G T A C C A C G C C A A G A A A A A A G A T G G Y A T T G 53280

L H G A L G F L L K Q F H G L L F P R G M P P H S K S P N P Q W F V N T L L Q R N 848
T T I A C C G G G C C C T A C C G G G T T T G C C T A A C A G C T T C A C G G C C T G T T A T T C C T C G G G G T A T G C C A C C A A C A G T T A A C C C C C A A C C C G C A G T G G T T T G G A C C C T G T T A C A C G C A C 53400

Q M P A D K L T H E E I T T I A A V K R F T E E Y A A I N F I N L P P T C I G E 888
C A G A T C C G C C G A G A T A A C T T A C A C G A G A G G A T T A C C A T T G C A C G C T T A A C C G G T T A C C G A G G A A T A G C A G C A A T A A C C T T A A T C A C C C C C A A C C T G C A T A C G G A 53520

L A Q F Y H A N L I L K Y C D H S Q Y L I N T L T S I I T G A R R P R D P S S V 928
T T I A G C C C A G T T T A T G G C A A T T C T T A A T A C T G C G C A T T C A C A G T A C C T T A A A C C T T A C T T C T A T A A T C A C C G G T G C C A G G C C C C C G T G A C C C A T C A C C G C T T 53640

L H W I R K D V T S A A D I E T O , A K A L L E K T E N L P E L W T T A F T S T H 968
T T I G C A T T G G A T T C G C A A G A T G T C A G C T C C G C C G G A C A T A G A A A C C C A A G C A A A G G C C T C T G A A A A A C G G A A A C T T C C G G A A T T A G G A C T A C C G C T T T A C T C A C T C A T 53760

L V R A A H N Q R P M V V L G I S I S K Y H G A A G N N R V F Q A G H N W S G L N 1008
T T A G T C C G C G G G C A T G A T C A C A G C T C C A T G T G C T T T A G G A A T A A G C T T A G G A T T A A T C A C G G A C C G G C G A G G A A A C C C G C G T T C A C G G A A T T G G A C G G G T T A A A C 53880

G G K N V C P L F T F D R T R R F I I A C P R G G F I C P V T G P S S G N R E T 1048
G G G G G T A A A A T G T A G C C C G C T A T T T C A T T G T G C G C A T G C C G C T T T A T A A T G C A T G T C C T A G G A G G G T T T A T C G C C C C G T A C A G G G C C C C G T G C G A G C A C A T A T G C C A T T G C A G C 54000

T L S D Q V R G I I V S G G A N V Q L A I Y A T V V R A V G A R A Q H M A F D D 1088
A C C T A T C C G C A C C A G T T C C G G G T A A T T G T G A T G G C C G G G C A T G T T C A A T T A G C C A C G C T G T G C C G C A G T G G G C C G T G C G A G C A C A T A T G C C A T T G C A G C 54120

M L S L T D D E F L A R D L E E L H D O Q I I Q T L E T P W T V E G A L E A V K I 1128
 TGGTTAACGAGCTTACAGCGATGAGTTTTAGCCAGACTTGAGGAGTACAGCACAGATTACCAAAACCCCTGGAAAGCCTGCTGACCGTAGAAGGCCTCTAGAACGAGTAAAGATT 54240
 L D E K T T A G D G E T P T N L A F N F D S C E P S H D T T S H Y L N I S G S N 1168
 CTAGATGAAAAAACAGACCGGGAGATGGGAAACCCCCACAAACCTAGCATTTAATTTGATCTTGTGAACCAAGCCATGACACACATCTAACGATTAAACATTTCAGGGTCAAAC 54360
 I S G S T V P G L K R P P E D D E L F D L S G I P I K H G N I T M E M I - 1204
 ATTTCAGGGTCAACTGTGCTCTTAAACGCCCGGAGACTCTTGTCTTAGTGGTTCCATAAAACATGGGACATACATGGAAATGATTAAACCTCCCTCT 54480
 TTATCCAATTAAAGCCCACACGGGGTGAGTGTACGTAATAAACAGTCATATTACATATTCTGTGTGTTCTTTTGTGTAGTCCTTACCCATATGACCTGTAATATAGTG 54600
 30 M E L D I N R T L L V L L G O V Y T Y I F Q V 23
 GTCTCCAACCATTCACTGAGTACAGTCAGTGGAGACTACAGCCGATAACATGGAAATGGATATTACGCAACTGTTGTTCTACTGGGTCAGTTACAGTACATCTTCAGGTG 54720
 E L L R R C D P R V A C R F L Y R L A A N C L T V R Y L L K L F L R G F N T Q L 63
 AACCTGCTACGTCGATGATCCAAGGGGGCGCTGCTCTTATATCGGTAGCGCTAACACTGTTGAGCTGTTTAAAGCTGTTCTCCGGGATTAAACCCACCTAA 54840
 K F G N T P T V C A L H A L C Y V K G E G E R L F E L L Q H F K T R F V Y G E 103
 ATTTGGAAAACACTCCACGGGTTGCTGCACTGGCATTATGTTATGTAAGGGAGAGGTGAGCCTGTTGAGTGTACACATTAAAACCGCTTGTGAGT 54960
 T K D S H N C I K D V F V S A F N L K T C O Y H H E L S L T T Y G G Y V S S E I Q 143
 CTAAAGACTCAAAGTGTCAAAAGATTCTTGTCTCAGCGTTAACCTAACATGCTACCGTGTGTTAACACATACCGAGGTTAGCTGAGTGAATTACAGT 55080
 F L H D I E N F L K Q L N Y C Y I I T S S R E A L N T L E T V T R F M T D T I G 183
 TTTACACGACATTGAGAATTTTAAACAGCTTAAATTACTGTCTATATTACGCTCTCGTGAAGGGCTAAACACATTGGAAACCGTGAACCGGTTTATGAGACAGATACTATAGGA 55200
 S G L I P P V E L F D P A H P C A I C F E E L C I T A N O G E T L H R R L L G C 223
 CGCGCTCAATAACCCACCGTGGATTGTTGATCCGGCCTACATGCTATGTTGAGAAATTATGTTAACAGCTAACCAAGGGAGACCTTACATGAGATTATTAGATGTA 55320
 I C D H V T K Q V R V N V D Y D D I I R C L P Y I P D V P D I K R Q S A Y E A L 263
 TCTCGCGATCGTACTAACGAAAGTCGGGTTAACGCTGGATGTTGAGCTATTGGTGTGTTACCATATACCTCTGATGTAACCGGATATCAAACGTCATCCGGTGAAGCGTAC 55440
 R T L O T K T V V N P M G A K N D T F D O T Y E I A S T M L D S Y N V F K P A P 303
 GAACACTCAAACAAAGCAGCTGAGTCAATCCATGGAGCAAAGACGATACTGGGAGAACATACGAAATTCGAGCACCATGCTGTTGATCTTAAATGTTTAAACCTGCCCCCTC 55560
 R C M Y A I S E L K F W L T S N S T E G P Q R T L D O V F Y D N H L D Y L N E H E K 343
 GGTGATGACCGCATACGGAGCTTAACGCTTAACTGGTAAACGCTTAACTCCACTGGAGGACCCAAACGACTTAAAGCTGTTGATAATTGGATTTAAACGAACATGAAAAC 55680
 H A E L T A V T V E L A L F G K T P I N F D R A F S E E L G S L D A I D S I L Y 383
 ACACGAGAACTTACACGGTAACGGTTGAGTGGCGTTATTGGAAAAACTCCACACTTGTATAGGGCGTTCTGAAGAACCTGGATCTGGATGCAATTGATAGTATTTCGGTT 55800
 G N R S S S P D S Q I E A L I K A C Y A H H L S S P L M R H I S N P S H O N E A 423
 GCAATCGCTCATCCTACAGAGCTAACGCTTAATTAACGCTTATGCTCATACGGCTCATCTGCGCCCTCATGGCTCACATCTAACCGAGTCAACGAGTCATGAAACGG 55920
 A L R Q L L E R V G C E D D L T K E A S D S A T A S E C D L H D D S S I T F A Y 463
 CCTTACGCCAACTTTAGAAAGAGTGGGTGAGGATGTTAACCAAAGGGCGAGTGACAGCGCTAACACATGGATGTCATGAACTGAGTACAGTAACTTTGCTTC 56040
 H G W E N L L S K A K I D A A E R K R V Y L E H L S K R S L T S L G R C I R E Q 503
 ATGGATGGAAAACCTGTTATCCTAACGAAATTGACCGCTGGAGAAAGAACGAGTATCTTGACATCTGCTTAACCGCCTCTGAGTATGTTAACCGAGTCAGCGAACAGC 56160
 R Q E L E K T L R V N V Y G E A L L Q T F V S M Q N G F G A R N V F L A K V S Q 543
 GCAAGAGCTAGAAAAACACTCAGGGTAAACGTTATGGAGGGCTTATGAGACATTGTTGAGTCAACCCCTAACACCGCTTATTTTACCGTGGAAACGTTGGGCTAAGGTTCCAGG 56280
 A G C I I D N R I Q E A A F D O A H R F I R N T L Y R H T V D A A M L P A L T H K 583
 CAGGGTGTATTATGCGACAATCCGATTCAGGAACGGCCATTGAGCATAGTTAACGAAATCCCTAGTGTACAGCTGCTCTAACCGCCTCGGTAGATGTTACCTGCACTAACATAAT 56400
 F F E L V N G P L F N H D E H R F A Q P P P N T A L F F T V E N V G L F P H L K E 623
 TTTTGAGTGGTCAACGGCCATTGTTAATCAGATGAAACCGCTTGTGACACCCCTAACACCGCTTATTTTACCGTGGAAACGTTGGCTATTTCGGCACTTAAAGAGG 56520
 E L A K F M G G V V G S N W L L S P F R G F Y C F S G V E G V T F A Q R L A W K 663
 AATGGCAAACTTTAGGGCGTGTGTTCAACTGGCTCTGAGCTTACGGCTTATGGGCTTTATGCTTCTGGGTAGAAGGGCTACTTTGCAACAGAGACTGGCTGAAAT 56640
 Y I R E L V F A T T L F T S V F H C G E V R L C R Y D R L G K D P R G C T S O P 703
 ATATTAGGGAGCTGTGTTGCAACACACACTTACCCCTGTTCTGAGCTGGGGAGGTGCGGTATGTCGCGTTGACCGTCAAGGATCCACGGGGTGCACGTCACCTA 56760
 K G I G S S H G P L D G I Y L T Y E E T C P L Y A I I Q S G E T G I D Q N T V Y 743
 AAGGTATAGGCACTTCCACGGACCTTAGACGGCATTTAACGTAACGAAAGAACATGCTCCCTTGCGCTTATTTCAAGTGGAGAAACAGGGATCAGCGAACATACCGTCA 56880
 I Y D S D V F S L L Y T L M Q R L A P D S T D P A F S - 770
 TCTACGATTCAAGCGTTTCTCTCTATACACCTTAATGCGCGCTGGCTCCGGATTCAACGGACCCGGCTTTCATAACCTCCGTTACGGGGTGTGTTATGCTTTATGCA 57000
 31 M F V T A V V V S V S P S S F Y E S L O V E P T O S E D I T R S A H L G D G D 38
 ATTTCTATGTTGTTACGGCGGTTGTCGGCTCTCAACGCTTATGAGAGTTAACAGTAGACCCACAACTACAGAGATAACCCGGCTGCTCATCTGGGCGATGGTCA 57120

E I R E A I H K S Q D A E T K P T F Y V C P P P T G S T I V R L E P T R T C P D 78
 TGAAATCAGAGAACGCTACACAAGTCCCAGGCCAAACAAAACCGACGGTTTACGCTGCCACCGCCAAACGGCTCCACATCGTACGATTAGAACCAACTGGACATGTCCGGA 57240
 Y H L G K N F T E G I A V V Y K E N I A A Y K F K A T V V Y K D V I V S T A M A 118
 TTATCACCTTGTTAAAAACTTACAGAGGGATTGCTGTTTATAAGAACATTGCGCTACAAGTTCAGGCTGTTAACAAAGATGTTACGCTGTTAGCACGGCTGGGC 57360
 G S S Y T O I T H R Y A D R V P I P V S E I T D T I D K F G K C S S K A T Y V R 158
 CGGAAGCTCTATACGCAAAATTACTAATGATAATGCCGATAGGGTACCAATTCCCGTTAGAGATCACGGACACCATGATAAGTGGCAAGTGTCTCAAAGCAACGTACG 57480
 N N H K V E A F N E D K N P Q D M P L I A S K Y N S V G S K A W H T T N D T Y M 198
 AAATAACACAAAGTGAAGCCCTTAATGAGGATAAAATCCACAGGATATGCCCTAATCGCATCAAATATAATTCTGIGGGATCAAAGCATGGCATACTACCAATGACAGTAC 57600
 V A G T P G T Y R T G T S V N C I I E E V E A R S I F P Y D S F G L S T G D I I 238
 GGTTGCCGGAACCCCGGAAACATAAGGACGGGACAGCTGGTGAATTGCGATCATTTGAGGAAGTGAAGCCAGTCAATACTCCCTTATGATAAGTGGACTTCCACGGGAGATAAT 57720
 Y M S P F F G L R D G A Y R E H S H Y A M D R F H Q F E G Y R Q R D L D T R A L 278
 ATACATGCCCCGTTTTGGCTACGGGATGGTCATACAGAGAACATTGCAATGGATCGTTTACCAAGTGTGTTAGAGCAAAAGGGATCTGACACTAGAGCATT 57840
 L E P A A R N F L V T P H L T V G W H W K P K R T E V C S L V K M R E V E D Y V 318
 ACTGGAACTCTGACGGGGAACTTTAGTCAGCCTCATTTAACGGTTGGAACTGGAAAGCCAAAAGGAAGGTTTACGCTGTTAGAGCTGGCTGAGGTTGAAGACGTAGT 57960
 R D E Y A H N F R F T M K T L S T T F I S E T H E F N L N Q I H L S Q C V K E E 358
 TCGCGATGAGTACGACAACTTTCCGTTAACATGAAAACACTTACACGGTTAACATGAAACAGAGTTAACATCTCAAGTGTAAAGGAGGAA 58080
 A R A I I N R I Y T T R Y N S S H V R T G D I Q T Y L A R G G F V V Y F Q P L L 398
 AGCCGGGGCTATTAAACGGGATCTACACAGGATACACTCATCTCATGTTAGAACCCGGGATATTCCAGCTACCTTGCCAGGGGGGTTGGTGTGTTCAACCCCTGCT 58200
 S N S L A R L Y L Q E L V R E N T N H S P Q K H P T R N T R S R R S V P Y E L R 438
 GAGCAATTCCCTCCCGCTCTATCTCAAGAATTGGCTCGTAAACACTAATCATTCACACAAAAACACCCGACTCGAAATACCGAGTCCCGACGAACGCTGCCAGTTGAGTGGC 58320
 A N R T I T T T S S V E F A M L Q F T Y D H I Q E H V N E M L A R I S S S W C Q 478
 TGCCAATAGAACATAACACCCTCATGGTGAATTGCTAGCTCCAGTTACATGACCAATTCAAGAGCTAACATGACGTTAGAACAGGAGTTACCTCGCTCGTGTGCA 58440
 L O N R E R A L W S G L F P I N P S A L A S T I L D O R V K A R I L G O V I S Y 518
 GCTACAAATCGCGAACCGCCCTTGGAGCGACTATTCCAAATTACCCAAAGTGTCTTAGCGGACCCATTITGGATCAACGCTGTTAACGCTGTTACCTCGCCACGTTACCTCGGT 58560
 S N C P E L G S O T R I I L O N S M R V S G S T T R C Y S R P L I S I V S L N G 558
 TTCTAATTGTCAGAAGTGGGATCAGATACCGCTTACCTCAAACCTATGAGGTTACTGTTAGACTACGCTGTTATAGCGCTCCATTAACTTAAAGTAGTTAGGTTAAATGG 58680
 S G T V E G O L G T O N E L I M S R D L L E P C V A N H K R Y F L F G H H Y V Y 598
 GTCCGGGACGGTGGAGGGCAGETTGAGACAGATAACGAGTTAATTGTCAGAGATCTGTTAGAACCTCGCTGGCTAACTAACAGGATATTITGGTACTACCGTATA 58800
 Y E D Y R Y V V R E I A V H D V G M I S T Y V D L N L T L L K D R E F N P L Q V Y 638
 TTATGAGGATTATCGTTACGCTCGTGAATCGCAGTGTGGGAATGATTGACTACGCTAGTGTGGATTTAACTTAACACTTAAAGTAGAGGTTAGCGCTGCAAGTATA 58920
 T R D E L R D T G L L D Y S E I Q R R N Q M H S L R F Y D I D K V V Q Y D S G T 678
 TACAAGAGACGAGCTGGGGATACAGGTTACTAGACTACAGTGAATTCAACCGCCAAATCAAATGCTTACCGCTGGCTTTATGACATGACAGAACGGTTGCAATATGATAGCGGAAC 59040
 A I M Q G M A Q F F Q G L G T A G Q A V G H V V L G A T G A L L S T V H G F T T 718
 GCCATTATCGCAGGCGATGGCTAGTTTCCAGGGACTTGGGACCCGGCCAGGGCTGGCTGGACATGTTGGCTTGGGACCCGGCTGGCTTACCGTACACGGATTACAC 59160
 F L S N P F G A L A V G L L Y L A G L V A A F F A Y R Y V L K L K T S P M K A L 758
 GTTTTATCTAACCCATTGGGGCATGGCTGGATTATGGTTGGCTGGACTGGTACCGGCTTTGGCTGGCTACCGTCTTAACTTAAACAGCCGATGAAGGCATT 59280
 Y P L T T K G L K Q L P E G M D P F A E K P N A T D T P T E I G D S Q N T E P 798
 ATATCCACTCACACCAAGGGTTAACAGTACCGGAGGAATGTTACCGGAGGAACACCAACCGTACTGATACTCCAAATAGAACAGGAAATTGGCAGTACACAAACACTGAC 59400
 S Y H S G F O P D K F R E A Q E M I K Y N T L V S A A E R Q E S K A R K K N K T 838
 GTCCGGAAATAGGGGTTGATCCGATAATTGAGAACGCCAGGAATGTTAACATGAGCTTACGCTGTTACGCTGAGGAAATCTAACGGCAGAAAGGGGGTTGTAAGCCAGGGGGTTGTTAATT 59520
 S A L L T S R L T G L A L R N R R G Y S R V R T E N V T G V - 868
 TAGCGCCCTTAACTTACGCTTACCGGCTTGCGAACCGGAGGACTCCGTGTTCCGACCGAGAATGTAACGGGGGTGTAAGCCAGGGGGTTGTTAATT 59640
 TTAATAAAATGTTAACCGTACTCATGTCCTACCGCATCACGGGGTTATACCCGATAATACAAACCGCTTGTGACCTCTACCGACCCGATACTAACCGGG 59760
 M E S S N I N A L Q O P S S I A H H P S K Q C A S S L N E T V K D S P P A I 38
 TTATTTATGGAACTGCTAACCTAACCGCTACACACCGCTCATCGCACATCCGCTAACAGCTGCCATCAAGTCTAACGTAAGAACAGTAAAGATTCTCCCCCGCGATT 59880
 Y E D R L E H T P V Q L P R D G T P R D V C S V G Q L T C R A C A T K P F R L N 78
 ATGAAGATAGGTTAGAACACACGCGGTAACACCGCCGAGCGTACACCCGAGACGTTGCTGTTGAGGACTAACCTGCTGAGCATGTCAGAACGAAACCTTTCGCTTAAAC 60000
 R D S Q Y D Y L N T C P G G R H I S L A L E I I T G R W V C I P R V F P D T P E 118
 GCGACAGCAATACGACTACTAACACATGTCAGGGGGCCGTCATATTCACTGCACTGGAGATTAAACGGGTGATGGGTTGATCCCGCTGTTCCGGATACCCAGGG 60120

E K W M A P Y I I P D R E Q P S S G D E D S D T D - 143
 AAAATGGATGGCGCCATAATTATTCAGGAGAACACCATCATCAGGGATGAAGATTCTGACACCGATAATTAACTTAACCTAACCCATAAAAACGCCTCT 60240

33 GTTTGTAAACACGACACCGCTAACAAAAAAACCAACACGCCCTTGTAACTGAAATGTAATACTTTATTTGGTTAACACGCCACCATCATGATTTGCAAATATCGC 60360
 - C R G . V W M Q N A F M D 594

GCGTCGCTGCCGTGGACCCCTGTATTAAGGGGCTTGGAACTCGCCTCCACTGCATTACATCTGTCACAGTGTATGTTGGGATGGCTGTTGATTTGGGATGACATAGA 60480
 ADD AT S G Q I L P A K S S A E V A N V D Q G V T D T H P T S T T N O S S C L 554

CCCCAACGCTTGAAGCTTGTAAATAAAATCGATATTGAGGATCCCGTGCCCTCTGGTATATTGATGGTGCACAAAGGCATTGTCGCCATTGGTATTTAGCTCTGTA 60500
 6 S V S Q L O K L L I S I R P D R T G E P I N T H H S L P M Q T G N O S K L E T 514

ACCTCTGTTGAGCTTGCACACCCCCGCAAGCTTCGCTGACCATAGAACCTCTGTCCTCCAAATATGATGGAGAACCTCGAGTCGGATGGCTATATACGTT 60720
 V E Q Q L K A V V G A L E E H Q G N S V R H R R Q W Y S P S V R R R B H T I Y T 474

GGTCACCGGAAATATATATTGAGGAAACTCTCGTCATTGAGACTCCCACATAAAAAGATCCAATTCCCTTGTATCCATGTCCTGAAATCCGTTCTGGACGACGG 60840
 P E G P F Y I N S P F E G D M Q S E G S Y F S D L E R Q D N S K F O R K G P R R 434

ACATCGGGTTGCTGAAATTTACACGGGGTCTGCAAGTCATAACCCCGTTCGGGCCATCGGTCTATAATCGGCACATTGCTTCCAAACGATTGGGTTGGGATATCCC 60960
 V D T K D P F N V C P T Q L D I G R E A A L A N M F A S M O N E L R N P P P Y G 394

GGAAACCGGAGCTGGTCCCGGAGGCAACCATAACCCCGTGTAGGTGGGAAGGGAGGCGGGGTGTGGAGATCCATATGCCGACGATATCTGTCCTTATTGGA 61080
 P F G Y P G G F H G P P C G Y G G T N P P F A P P P H P S G Y P G V I Y Q G H N P 354

GCTCAATTGATACTCGGGATTITAGTCGCCCCGTTAACAGCTGTAAATAACCGGGTAGGTACAGTACAATTCCCTCGGGTGGAACCGCCGACGGGGCTGTGGT 61200
 A G I S Y Q P N K T Q G T L L Q S Y Y A T P I L V F E G G T P V G S P P Q P S I 314

TTACTACGGTACCTGCTACAGAGCCATACGCTGTGCTTCTACACAACCTGCAACCTTTAAATCGGGAACAGTCCTTACAACTTCATTTGATCCCCAACCCGAGA 61320
 N S A N G A V S A M D S O N R C L Q T Y K L H P V T K E C D E N S E G V G L A L 274

TCGTATATGGGGCCCGGGGGTGGATGTGGCTTATAACACCCGGTGGTAATTAGACTCCACCCATTAACTGGGATTCGAGTCATATCCGGTGTAGCTGAGA 61440
 D Y I P G G G P H F T A N I V G A N P L K S E V G N Y N T I R A L G Y G T S A Q L 234

TAACCGTACCCATAATTCCGCCCTCCGCTACGGTTGCAACACGTCCTACATACTCTAAAGCATATTGTCACGGCTGTGGATAATAACACCTTGGCAGTTATTCGCTA 61560
 Y V H G M J G A E R R R K A V Y D W R D R L L M N N Y A T S L L V K A L K D E S 194

ACCTCCATACTTTATTAAACCCCGTAGCTTAAACAGGCAATAACCCGGCTTACTTTCCATCGATAAAACCCGGAATGGTCAATTGAGATTCCGGGTACAGTCATAATTG 61680
 V K V V K N L G A Y D K V L S L L R A K S E N S L V R F P E I S S E P T C D Y N 154

ACACTGTCACCGCTTCACACACATAACCGAACATGGTAAAAAAATACCGTCTGGTATCTACCGGGACAATCGTTGAGACAGGGATACGGAGGTAAGTAAATTG 61800
 V V T G V R R G V V C L A V H T F F N G D P I E N P S L R K S S L S V S P L Y N 114

GTGACCAAGTATAACCGCAGCTCTAGCGGAGATAACAGAATCTCTATTCCAAAAAAATCGATGGCCCTCAACAGCAGGGCATGTTGAGGGCATCTAACGATAACCAA 61920
 T V L Y L A R E L P S L V S D R N G F F N S H A A E F L Y A H L Q P C R V I G L 74

AAAAAAGTCCCGTAGTCCTCAATGATTGCGATTACTTCACCCACGACACAGTCCTTCGATGATCGATGTTATTGGTATTACTAGTAGGGGCAAAGCCGACCAATCTC 62040
 F F P G R I D E I X I A I V E G V Y C D K R H D I N I P I K S T P P L A S R V I E 34

GGGTAATTAAATTCCCTCGCTTGAATAAGGTAATACCGCCAGTAACTGCTTCACAGTCCTCTCGTCAGCTCACAGGCTATAACACCCACGGACCGGA 62160
 P T I N L E G E D K S V L A L Y G A V Y L A E C N E E D A E A A M 1

TAGTAATCTACGGTGGAGGCAACTGGAAATGACACCCAAACAGACAAATATAGAAGATCATAGTCAGTCACTGTTAACCTGAACTCGCAAGGGGACTTTCTCCATGGC 62280
 34 - P T S A F Q P I V G L C Y F Y L L D Y D S H V N F Q A L R R S E E L A A 544

CCCTACACGGGTTGGCATTAACTCCAGTCCTCCGTTCAATTCAGCAACATAACACTGCTGATGACACCGGGCTATGGGATGGGATGTAATTGTCGGGTGTT 62400
 R V R P Q H M L M G L G R E Y O L M V S D H I S V Y A A I P L S T V N Q R R H E 504

CTAATTCCAAATGCAATTAGCTTATGAGCCGATCTGGTACTGCCAGAAGAAATATCTTACCGGTTCTCCATAACTCCACGACTAACGCTGGGATGCGCTCAAACAAAGCCAA 62520
 L E L A I L S I L R D Q Y Q G S S I D I Y T R G L S G R S L Q T H A D L C L A V 464

CTAATCCAGGAAACATTCAGTCAGCTCTGGTCCGTTAACGCTACAGTCAGTCATATATCGTTCACATAAAATGAAAGTTATTACCGCTTAAACTCCCATCAAACC 62640
 L G P F M E T L E T T R N L T Y L P A I Y R E C L F Q F N N N G S K F K G D F G 424

CCGTCGCTCCGGCAAGGTTACCTGGTAGGGGTTCTGCTGCTGACACATCAAACCCAGTGTGAAATTATTTAGTTATCTCGTACGTTCCGCTCAAATAGC 62760
 T A G R L I V N N T P T G T A E S V I L G L Q F N N K L K D G Y V N G N W L L R 384

GCCTTAATAATAACCGCGTAATGCTGTAATTGTTACCTTAATAGGTTGGCTTCCATAAGAAACAGCTTGGCCGCTCTAAATACGCCGCCGCCGCTGTGAATCTGT 62880
 R L L L A T I T D I T L R L L T O D E N L F Y N Q A R E L Y A A A A Q Q I K D 344

CCACATATGGGATGTTGGATCAATAATGCTTAACCCAGGTTAACCTGTCAGTCAGGTCAGGGTAGGACCTGCAACCGTATAAGGCACTCAAACAGAATGTCAGCTAAGG 63000
 V Y A T H N R D I I O N V G P N F Q G P A P P L V Q L R I L A D M V S H S T L A 304

CCCTTGATCATACGCCCGCACGGCATGAAACTGGTCGGTAGACGATCATAGCAAAATTGATACTGTTTATTCGTCGTTGTCATAATTACACAAATGTCAGTAT 63120
 6 Q D Y R G G R A H F Q D R P L R D Y C F Q Y S N K N E H T M Y N V F T E T Y 264

143 ATTCCGGTAGGTGCTCTATAAGTTCCCGAAGGACGAACTTGAGGTTCTGGACACTATTAGATGTCCTATACATTAATATAACATAATACCGCACACTCGAACGCCAGTACGCTC 63240
 60240 EPLHEILNGFSSVQPEHVSNSTRYMLYLCLVACEFASYAR 224
 G 60360 ATATCCTCCAACTACATTCCTCCCGCCGACTGTAGACATGTTACCGGGCTTCAATACACGCGGGAAATGCCGCGCTTCAATACACTCCGCTGCAAGCTACGGGCCATCTAACA 63360
 594 DGYYNYMRGASQLCTVTTHNMFTRSIRGDOKCDVRAAVPRLV 184
 A 60480 CAAAGCCGTCCTCGAGAGTACGATACCATGGCCGAAACAACTCCCTGGAGGTTATGGCCCTTCCTCCAAAGTACACCCAGGGTATATAAACTTCACTGAAAGTTGATCGTACT 63480
 554 LREOLTRYWPGFVIGPSNNGRARGLYVLTIFDVFQFNTDYQ 144
 A 60600 GCAACGGTCATCATTTGGCAATCTGACCTCGGGGTATAGACTCATTCGTTATTTCTGACGTYGACATTCCTCAGATTGTCATCTGCTTCGCCCTGCCAGCG 63600
 514 LPADNKAIQVEPHISENRRIETRTCEESQADAEEAEAAAT 104
 F 60720 TCTCCAGGAAATCCAAACCTTGGCCATGCGCTGAGTTGTTCTCGAGGGCTTAAACGACGATCTATTCCGTTGGTACGTAATCGTTCCCGCGAAGGGTGTCTAATGCCGAA 63720
 474 ELSDLVKANRTLQEEELPKLRRDIETPLTITE6RLNOLAAY 64
 I 60840 CGGGCCGCGATTTTAAACGTTAACGTTAACGTTAACGCGGATTCTACGCCCTCTAACCTAACACGGGGAGCCGCCAGTAGTGTATGGGAAAGTTGGGGCTATAAAGTCT 63840
 434 AAANKLTLTNKELDOPNMRGRLEFAPATWYHIPFNPAPIFNK 24
 D 60960 TAGGGTACAGAAAATCCCACATTTTCGGAAACGAGATAGTCGAACCCATATCTCGCGCTATGGTCTGCAAGAACAAAGTCAACTGGCGTAATATAACCGGTACTGC 63960
 394 TTSLFIGCKNPFSISGFYRATM 1
 61080 TTTAAAAGCTGTTTCTTACCCATGGGAAACATCCCGTTACTTGTAAATTCACCAAGCCTTAAGAAGGCCCTCAAGGGGAAATCCACCCACAAGCTGCTTTCTT 64080
 354 - GHSFNGTITISQLIIGGGCAGLSPRRLPLDVGCAANE 225
 61200 CAAACTTGTAAAGCGGAACGATGGCATGTTGCGACGCTTTCGCAAGAGAACATACGTAATTTCTTTCGATAGACGCTCTCGCTCTAACGGACCTTATGGGGGTAT 64200
 314 FKLTLASRHCSKARKKALSCYHIKKKCLRRRERVSRIPPPTY 185
 61320 ATTCCGGTACATCTCCAAATCGGAGCTAGCATAACAAAGGTTCCATGAATCACCTTGGGGTAACCGAGTTACGTTAACAGGGTCAAGCCCGTGGAGATAAACACAAAGGG 64320
 274 EAIVNHLHSALHVLNGHIVVKPPLRTVOLNLGRQSVFVLLP 145
 61440 >>> 5' end of dPyk mRNA ***
 234 GGGTCACCAATTTCATCAGATCCCGTGGGTGTTCTTAAAGCCATGGTATCCCTCAGCTGGCATAACCTCTGGAAACTGGTATGTTAGGGGTATGTTATATTAG 64440
 1
 61560 TTYMIEDSGTPTEKILANTDRLQRMGEFCQHYKTPTHIN 105
 194 CGCTAAACGGCAAGATTAACTTACATAAACAAACGGTTCTCCGCACACTGGATTCCGTTGATAATACAAACAAACATCGGGCGTGGCTCCCAAATTAACCTCAAACG 64560
 SFRCSKLEVIFCVTKGAGSSSETQIIICVCDP R R R G L N V E F S 65
 61680 ***
 154 ACATTGATATGGTACAGCCCTTGAACATCCACGTTGGGATAACGGCGACAGGGTTGGCCAGCCTGGGTGAACGGCTCCGGAACACTGACGTTACATATCTTTTG 64680
 61800 M S I R V A R Q V D Y H S L P S L L K A L R P N F A D A F G R R V N D I O K K L 25
 114
 61920 GTACATCGAAAAACGAGTGGCAACGTTGCTCCAAACGAAACACTGGCCGAAATTCGACTAGCGACATATTGAGTTCCGTCAGAAGATAACCTAACGCGTTGCTACA 64800
 74 YD Y F R T H C R Q G L R F C K A R I R S A M 1
 62040 H S T D K T O V K N G V L R I Y L D G A Y G I G K T T A A E E F L H H F A I 38
 34 ATAAACATGTCACGGATAACCGATGTTAAATGGGCTTGGCTGTTGGGAAATTCGACTGGGAAACACTGGCCGAAATTCGACTGGGAAACACTGGCCGCAAGAATTTTACACCACTTGTCA 64920
 62160 T P H R I L L I G E P L S Y W R N L A G E D A I C G I Y G T Q T R R L N G D V S 78
 1 ACACCAACCGGATCTTACTCATGGGAGCCCTGCTGTTGGGAACTTCGAGGGAGGACGCCATTGGGAAATTACGGAACACAAACTGGCCGCTTAAAGGAGCTTGG 65040
 62280 P E D A Q R L T A H F Q S L F C S P H A I M H A K I S A L M D T S T S D L V Q Y 118
 544 CCTGAAAGGCCAACGCCCTACGGCTCTTTCAGACGCTGTTCTCCGATCAATTGCTGCAAGCTGGGAAATCTCGGCTTGTGACGAAACAGTACATCGGATCTCGTACAAGTA 65160
 62400 N K E P Y K I M L S D R H P I A S T I C F P L S R Y L V G D M S P A A L P G L L 158
 504 AAAAAGGCGCTATAAATTATGTTATCCGACCCACACCAACTGGCTCAACTATATGTTCTCCCTGAGATACTTGTGGGAGATATGTCCTCCAGGCTTCTGGTTATG 65280
 62520 F T L P A E P P P G T N L V V C T V S L P S H L S R Y S K R A R P G E T V N L P F 198
 464 TTACGCTTCCGCTGAACCCCCGGGACCAACTGGTAGTTGACCGTTCACTCCAGTCATTATCCAGAGTAAGCAACGGGGAGACCCGGAGAAACGGTTATCTGGCTT 65400
 XbaI :
 62640 V H V L R N Y Y I M L I N T I I F L K T H N N H A G W N T L S F C N D V F K Q K 238
 424 GTATGGTCTGAGAAATGTTATATAACATGCTTAAATACAACTTAAACTAACACTGGCAGCGGGCTGGAACACACTGCTATTTGTAATGATGTTAAACAGAAA 65520
 62760 L Q E S C I K L R E V P G I E D T L F A V L K L P E L C G E F G N I L P L W A 278
 384 TTACAAAATCCGAGTGTATAAAACTACGGGAAGTACCTGGATTGAAGACACGTTATCGCCGCTGCTAAACTTCGGAGCTTGGGAGATTGGAAATATCTGGCTTATGGGCA 65640
 62880 W G H E T L S H C S R S M S P F V L S L E Q T P Q H A A Q E L K T L L P Q N T P 318
 344 TGGGGATGGAGACCCCTCAACACTGCTACGAAGCATGTCCTCGCTGTTACGTTAGAACAGACACCCAGCATCGGGCACAAGAACACTCTGCTACCCAGATGGCCG 65760
 63000 A N M S S G A M H I L K E L V N A V Q D N T S - * 341
 .304 GCAAAATGCTCCCGGTCAGGAAATATTGAAAGACGTTGTTAATGCGCTAGGACAACACTCTCTAAATACCTAGTATTACGTTAGTACCAAGTAAAGATGATAACATG 65880
 63120 >>> 3' end of dPyk mRNA
 264 TCATACTGGCTGAGCTGTTCTTATATGCGTCTTAACTACATCTTAAACGGAGTGGTAATATTTAAAGCCGCAAGCTGTTG 66000

37 M F A L V L A Y V I L P L W T T 16
 GTGGGTGAGGAGGGGAAAGGCACGCTGTGATAACGTTGCCGTATTTGAGCGAACGACTATGTTGCCCTAGTTTACGGGTGTAATTCTTCCTTGGACAC 66120

A N K S Y V T P T P A T R S I G H N S A L L R E Y S D R N M S L K L E A F Y P T 56
GGCTAATAATCTACGTAACACCAACCCCTCGCAGCTCGCTATCGGACATATGTCGCTCTACGAGAATATTCCGACCGTAAATGTCCTGAAATTAGAACGCTTATCCTAC 66240

G F D E E L I K S L H W G N D R K H V F L V I Y K V H P T T H E G D V G L V I F 96
 TGGTTTCGATGAAGAAGTCACTAAATCACTGGGAAATGATAGAAAACACGTTCTGGTTATGTTAAGGTTACCCCTACACACAGAACGGAGCGCTGGGTATATT 66350

P K Y L L S P Y H F K A E H R A P F P A G R F G F L S H P V T P D V S F F D S S 136
 TCCAAATAACTTGATTCGCCCCATCCATTCAAAGCAGAACATCGAGCACCGTTCTGGACGTTGGATTCTAGTACCCCTGACACCCGACGCTGAGCTTCTTGACAGTC 66480

F A P Y L T T O H L V A F T T F P P H P L V W H L E R A E T A A T A E R P F G V 176
 GTTTGCCTGATTTAACTACGCAACATCTGTCGTTACTACGTTCCACCAACCCCTGTATGGCATTGGAAAGGCTGAGACCCGACACTGCAAGAACGGCTTGGGT 66600

S L L P A R P T V P K N T I L E H K A H F A T W D A L A R H T F F S A E A I I T 216
 AAGCTTTTACCGCCTCGCCCAACAGCTCCCAAGAATCTACTCTGGGATTCGCAATTCGATACTGGCCACCCGTTGGCTCTCACAAACGACTCGGGCTGTCGAGTAATTTGGGT 66720

N S T L R I H V P L F G S V W P I R Y W A T G S V L L T S D S G R V E V N I G V 256
 CAACTCAACCTGAGAAATACAGTCCCTTTGGCTGGATTCGCAATTCGATACTGGCCACCCGTTGGCTCTCACAAACGACTCGGGCTGTCGAGTAATTTGGGT 66840

G F M S S L I S L S S G P P Z E L I V Y P H T V K L N A V T S D T T W F Q L N P 296
 AGGATTATGAGCTCGCTCATTTCTTATCTCTGACGACCCGATAGAATTATGTTGACCATACAGTAAACGACGGTTAACAGCACACATGGTCCGCTAATCC 66960

P G P D P G P S Y R V Y L L G R G L D M N F S K H A T V D I C A Y P E E S L D Y 336
 ACCGGGCTGGATCCGGGACATCTGGGATCTTATCGAGTTTACTTGAGCTGGTTGGATATGAATTTCAGTCGATGATATGGCATATCCGAAGAGAGTTGGATT 67080

R Y H L S M A H T E A L R M T T K A D Q H D I N E E S Y Y H I A A R I A T S I F 376
 CCGCTATCATTATCATGGCCACACGGAGCTGCGGATGACAACGAGGGATCACATGACATAACGAGGAACCTTACCATATCGCCGAAGAATACCCACATCAATT 67200

A L S E M G R T T E Y F L L D E I V D V Q Y O L K F L N H Y I L M R I G A G A H P 416
 TGCGTTGCGGAATGGCCGATCACAGAATTTCTGTTAGATGAGATCGTAGATGTTCAATTAAATCTTCAATTACATTTATGCGGATAGGAGCAGGAGCTCATCC 67320

N T I S G T S D L I F A D P S Q L H D E L S L L F G Q V K P A N V D Y F I S Y D 456
 CAACACTATACCGGAACTCGGATCTGATCTTGCCGATCACGTTCTGAGCTTACACTTCTGGTCAAGTAAACCCGAAATGTCGATTATTCATGAGCAATTT 67440

E A R D Q L K T A Y A L S R G Q D H V N A L S L A R R V I M S I Y K G L L V K Q 496
 TGAGGCCGTGATCAACTAAAGCCGATACCGCTTCCGTTGGTCAAGACCATGTAATGCACTTCTCGCAGCGTGTATAATGAGCATATAACAGGGCTGTCGAGCA 67560

N L H A T E R Q A L F F A S H I L L N F R E G L E N S S R V L D G R T T L L N 536
 AAATTAAATGCTACAGAGGGAGCTTATTTTGCTCAATGATTAAATTCGGGAAGGACTAGAAAATTCATCGGGATTAGACGGTGCACAACTTGTCTTAAAT 67680

T S N C T A A H A T Q A A L N I Q E G L A Y L H P S K H M F T I P N V Y S P C M 576
 GACATCATGTCAGGAGCTACGGCCACGGCAAGCAGACTTAACATAAGAAGGCCGACTTAAACATGTTACCATGTTACAGCATGTTACACATGGTCAAGTC 67800

G S L R T D L T E E I H V M H L L S A I P T R P G L N E V L H T Q L D E S E I F 616
 GGGTCCCTCGTACACCTCAGGAAAGGATTCTGTTGATGTTACCTCCCTGGCAATACCGCCAGGACTTAACGAGGTTATGCAACAGTCAGCCGCGCGTAACGGAGA 67920

D A A F K T M M I F T T W T A K O L H I L H T H Y P E V F T C Q D A A A R N G E 656
 CGACGGGCACTTAAACCATGATGTTTACCATGGACTGCAAAGATTGCAATACTCCACACCCATGTTACAGAGTATTCAGTGTCAAGATGCGAGCTGTC 68040

Y V L I L P A V Q G H S Y V I T R N K P Q R G L V Y S L A D V D V Y W P I S V V 696
 ATATGIGCTCATCTCCAGCTGTCAGGACACAGTTATGATTACGAAACAAACCTCAAAGGGTTGGATATTCCTGGCAGATGCGATGTTATAACCCATATCCCTGT 68160

Y L S R D T C V S E H G V I E T V A L P H P O N L K E C L Y C G S V F L R Y L T 736
 TTATTAAGCAGGAGACTCTGGCTGTCGACATGGTGTATAGAGACGGTCACTGCCCATCCGACAATTAAAGAATGTTGTTATGCGGAAGTGTCTTAGGATCTAAC 68280

T G A I M D I I I I D S K D T E R Q L A A M G N S T I P P F N P D M H G D D S K 776
 CACGGGGGGGATATTGGGATAATAATTATGACAGAACAGAACAGAACACTACGGGCTATGGGAAACTCCACATCCACAGACATGGGGGATGACTCTAA 68400

A V L L F P N G T V V T L L G F E R R Q A I R M S G O Y L G A S L G G A F L A V 816
 GGCCTGTTGTTGTTCTAACCGAACACTGTTGCAACGACGACGACAGAACATGCGGACAATACCTGGGCTTCTGGAGGGGGCTTCTGGGG 68520

V G F G I I G W M L C G N S R L R E Y N K I P L T - * 841
 AGTGGGGTTGGTATTATCGGATGGATTTGTTGAAATCCGGCTTGGAGAATATAATAAAACCTCTGACATAAAACATGTTATAAAAGTCACTATAACGCTCCAGC 68640

38 CAAACTTATTCGCGAATAACACACTACCTTGGGTTTTCCGCTCCAAATGGTGTGACTCTACCAAAAAAATAGCGCCTAAATGCTATATAACGCTCCAGC 68760

* - R Q T K K G D G F P T Q H V R G F F L A G L Y A I V R R G A 512

AAAATACGGTCAAGGCAATTACCGGATATGTTGAGTACAGGGCAATGGGATGGTGTACCCATAAACGGCATAGCGCACAGCGCCGTTATGCGGGGCTCAGACAG 68880

F Y P E F A N G S I T N Y Y L A I P I S S G L L R C L R V A G N Y C P D G S Y L 472

GGTATCTAGTACCGGATATCTCATCTGCTTCCGAGAGAACATCAACCGAACAGTATCCGATAACCAACTCTGTTTCAAGGGCTAAATTGCAAGGGCTTACAGTCAG 69000

T D L Y R S I E Y E H R E T V S V D V P V T D S L G V G T K A L R L I R V G E K 432

TTTGCAAGATGTGACGTTCTTGTAAAGGGAAAGCTGGGGAGTGGTAAGAACACAAAGTTCAGCCAAGCTGCCATAAACCCACTTCCCTCAAGGGCTGTTGCTGTATCCAC 69120
 KALHSTEKVCPLQPLPLFLLTTEALTGIFGVYERLLSNATD 392
 AATGGTCGTATTAAATCTTGAGCCAACCTGATCGTGTACATCACTGGTAACGCCATAACACCGCTTAAATCTTCATAAACGCCATAAACATTAAACGCCCTCCACCGAGA 69240
 ITRILDOAVQDODSPLANVYSRTLDEIVAYCNFAEWRS 352
 CAGTATATTTGACAATCACGAAACCGTGTACAGGAGCTCAGTACAGTCGCTAACGATCGCCGAAAAAATGGAAAGTTCGCCCCCTGGCCATAACGCCAACAATGCACTGAGTATATATCGT 69360
 LIYQVIVFRCSTLDRGYAHGFFPLEGREGYVAVVATYIT 312
 CTCACGGGCTTCATTAAGTTCATTCAGTCAGGCCATTTCGCTTTAAATACCTCGTCCGCAAAAAAAACCCACATGATAACGCCGGGATAACATGAGTAGTGGCTTATG 69480
 ERAENLEDELGPWKRRAKFIYEDAFFVACSLARICHTTAKH 272
 GCGAGGATCCAAATGTCCTACCCGGGGATGGCTCAATCTGTACAAAGTACTTAGTGTAAATATGATCGGACTTACCCGCTGAGCTGTTCTCAGAAATACGCCCTACCCGA 69600
 RPDWIDMWVRPITRIOVFNSLTIHDSSKKRRRPQKESYPEGS 232
 AATCCGCACATCATCTGCTTACGTCCTCCGTAACACATACAGCGCCGCGACTAACATTAACTATGTTTCTCATCGCTGTTACTCCGTTAACGCCGCTCGTATCCGGCGT 69720
 IPVDOAKVDETUVDAARRSVIISTKEDDNSGHLADRIEPT 192
 CCCGTCGAATAATCCACTCACTAGCTCTGCAAACCTTCGTAACCTCAACATACGCATATAACCAATGAAAACCGCTTCGTTGGTAGCTACATAAGCCATTGGTATTAAAT 69840
 GDFLGSVLEOLSEPLELMRMYVL SFSAEENPVYNFGNTHI 152
 GCGGGTGGGGTGGAAACAATTTCAGTTCGCGCGTAACATCACCCCGCCACCAATGTTAACATGCGTACGGGGAGGGGACAGAGATACTGCGAGCTGAGGTCTCCAC 69960
 ATPTPPFLKLKNERTYDVGAVLTLHTVPLSVALYDALTPDEV 112
 TTCAACATCAATGTCGCAAAGGTGCGGATCCACCCGCCCCGATCCGCTGCAAGTAAGGCACTCGCATCCAAAACACGCACTTATGGATGATAACGCCCATGCTTCCGGT 70080
 EVDFTGCLDRDVOAGSGAALLAVRDLFVCNNHNSVAWTKG 72
 CGGATTAAGCTACTTCGAAAGCTGCAACAGATCTGTCACCGGGTCCGACCCCGGAAACACATGCTCAAATTTCGCTCGCCCTCATGTTGGGGTATGGAAATTAAAGCA 70200
 RNLESRVYSAVSRDGPVGVRVHVGFKAIEEGGHKRPISILC 32
 TCCCCCGCATATAAAATACGCCCTGGTAGCACGCTCGTTAAATAAACGTTACCCGTTATAAGATACGGTTGAATGATGATGAAATCCATATAAACGTTATCGAACATTAACC 70320
 GGCIFFYARTARENFYFTVGNHYSVTSHYPFEM 1
 TCGAACTTGGCTCCCGTGTGATGTCGCAACCTTAGGTCACCGAATATGAAAATATAACTACGCCAACATTCAAAACCCGTTGATCATTAACTGATATGAAAAG 70440
 ATCCAATCAATCAACCACACCTCTACCGTTAGCACGTCAGCTATGTGACATGCTCCAAACATGAAACATTAGAGGGGTTATAACAGTGTCAAGGGGGTATATT 70560
 MNPPQARVSEQTKDLL 16
 39 CTACATAATCAAGGATCGGCTTAACTTGTCAACATTTCAGTACTTGGACTATAACCTGGACTGAACTGTTATGAAACCCACCCCAAGCCCGCTCGGAACAGAACAGACTTGCCT 70680
 SYNVNOHPEEDAKVCKSSDNSPLYNTHNVMLSYGGDTDL 56
 AGCGTTATGGTTAACAGCACCCGAAGAGGAGCACGAAAGTGTGTAATCCAGTATAATTACCCGCTTATAACACCATGGTTATGTTATGCTATGGGGTATACGGACTTACTATT 70800
 SSACTRTSTVHRSASFTOHSVYIISTVLIQPIICCIFFFFY 96
 AGCTCTGCTGATGACCGCAGTACCCGTAACAGGTCGGCTTACGCAACACTCCGTTATATTATATGTTACCCGTTGTTACCAACAAATATGTTGATCTCTTTTTTAC 70920
 YKATRCMULFLTAGLLTILHHFRLIINLLCVYRNIRSDSL 136
 TATAAAGGACACGGCTGTGCTTATCACAGCGGGTTACTCTGACGATCTACATCACTTGCACTTATTATGTTACCCGTTGTTACAGAAATACGATCAGACCTGCTA 71040
 PLSTSQQLLGLIIIVVYTRTNLFLCITAAYVTLFIDTRYFFLIT 176
 CCCTTATCTACATCCCAGCAACTGCTCTGGAAATTGTTGACTCGAACATGCTTATTTGTTACCCGTTGTTACCTTCTTATAGACACCCGGTGTCTTTGATTAC 71160
 GHLOSEVIFPDSVSKILPVSNWPSPAVLLVMAAVIYAMD 216
 GGACACTTGCACAGTGAGGTTATTTCCAGATAGCTGGTCAAAATCTCTGCTGCTGGGGTCCAAAGTCAGCGCTGTTACTGGTAATGGCGCAGTTACCGCTATGGACTGT 71280
 LVDTVFSFIGPRVWVRYMLKTSISF - 240
 TTGGATGGACACGGTATCCTTATGGCCAAGGGTGTGGGCTGGTTATGTTAAACATCTATTCGTTAGTCATCAATAATGTAATGTTGCTAAACAAATAATG 71400
 TTGGGTTATTTATAATTACCGCCCCCGTGTACTGGAAACCCCATACATATGTTCACTCATCAAACCTCTCGCAGTTCTGTTCCGACACGTTACAGTCCGGATTCAAG 71520
 MTTVSCPANVITTTESDRIAGLFNIPAGIIPPTGN 34
 40 TCGCAACGGCTGTGACAAAATGACAACGGTTCTGTCCTGCTAACGTTACTACAAACGGTATCTGCTGTTATGCTGGGGTTATTAACATCCAGCGGGGATCATCCAACTGGAAA 71640
 VLSТИEVCAHRCIIFDFFKQIRSDODNSLYSAQFDIILLGTYC 74
 TGTGCTGCTAACCATAGAGGGTGTGTCACACCGTTGCTTATTTGTTAAACAAATACGATCAGATGATAACAGCTTACTCGGCTCAATTCGATATCTTTGGGGACATCTG 71760
 NTLNFVRFLELGLSVACICICTKFPPELAYVRO6V1QFEEVQOP 114
 CAACATCAATTGTTGCGTTCTGAAACTGGACTGCTGCTGCTGCTGACTAAATTCCGGAGCTGGCTTAGCTGCGAGATGGCGTTATTCATTTGGGGACATCTG 71880
 MIARDGPHPYDQPVHNHYMVKRHKRSLSAAFIAASEALSL 154
 CATGATAGCACGCTGATGGCCACATCCCGTCGATCACGCTGTTACATATTATGTTAAACGCGGATACACACCGCTCGTAAGCCGCTGTTGCAATGCACTGGAAAGGGTGTGAGTT 72000
 LSNTYVODGETEIDSSLRIRAIQOMARNLRTVLDLSFERGTA 194
 GTTAAGTAACATATGTCGATGGGACAGAGATTGACTCATCGTTACGTTACAGGCTATCCAAAGATGGCTCGTAATTAACGACCGTTGGACTCATTTGAACGAGGCACTGCCGA 72120

Q L L G V L L E K A P P L S L L S P I N K F Q P E G H L N R V A R A . A L L S D L 234
T C A C A T T C T G G T G T C T A T G G G A G A A G C C C C A C C G G T A T C G C T G C T T C A C C A T T A A T A A T T C C A A C C G G G G G A C T C T A A T C T G T G C A C G C G G C C C T A C T T C G G A C C T 72240

K R R V C A D M F F N T R H A R E P R L I S A Y L S D N V S C T Q P S V M V S R 274
C A A C A G T G A G A G T C T G C G G A T G T G T T T A G G C C G A C C G C C A G G G A C C T G G T G A T C T C T G C G G T A T C T G C G G A T A T G G T T C T G T G C A C C C A C C A T C G G T G A T G T A T C A G C G 72360

I T H T N T R G R Q V D G V L V T T A T L K R O L L O G I L Q I D O T A A D V P 314
A A T A A C T C A T A C A A C A C T C G C G G A C G G C A G G G T G A C G G T G T G T G T A C A A C A G C A A C C T T A A A C G C C A C T T A C A G G G A I L Q I D O T A A D V P 314
A G T A A C A T A C G C G G A A A T G G T C T A C A C G G G A C A A C T T G G T A C C G C C C T T G T G A T G G G A A A G G C C T C C C G G A A T G G T G A T G A T G A C C C C G C C A T C T C T G T A T A A C C G G C C T A A 72480

V T Y G E M Y L Q G T N L V T A L V M G K A V R G M D D V A R H L L D I T D P N 354
A G T A A C A T A C G C G G A A A T G G T C T A C A C G G G A C A A C T T G G T A C C G C C C T T G T G A T G G G A A A G G C C T C C C G G A A T G G T G A T G A C C C C G C C A T C T C T G T A T A A C C G G C C T A A 72600

T L H I P S I P P Q S N S D S T T A G L P V N A R V P A D L V I V G D K L V F L 394
C A C G T T A A C A T A C C G C T A T A C C C C A C A T C C A C T C C G A T T C A C G C A C G C T G G C T C C G G T T A C C G C C G T G T T C T C G C G A T T T A G T G A T T G G G G A T A A C T T G T A T T G 72720

E A L E R R V Y Q A T R V A Y P L I G H N I D I T F I M P M G Y F Q A M S M D R Y 434
A G A G C A T T A G A A C G C G G G G T C T A C C A A G C T A C C G C G T G C C T A C C C C T T A T T G G A A A T A T A G A T T A C G T T A T C A T G C C A T A G G G A G T G T T C A G G C A A A C T C C A T G G A C A G A T A 72840

T R H A G G O F S T V S E O Q D P R Q F P P Q O 6 I F F Y N K D G I L T O L T L R D A 474
T A C A C G A C A C G C C G G C G A T T T C A C T G T A T C C G A A C A G G A T C C A C G T C A T T T C C A C C C C A A C G G G A T T T T T T T A T A A A A G A T G G G A T T A A C A C A G T G A C T C T C G T G A T G C 72960

M G T I C H S S L L D V E A T L V A L R Q Q H L D R Q C Y F G V Y V A E G T E D 514
A A T G G G T A C C A T T C G C C A C A G T G T C T G C G G G G C A C T T G T G C C C T C C G C C A A C A C T T A G T G C T A G T G T T T T G G T G T A T A C G G C G G G A G G G T A C A G A G G A 73080

T L D V Q M G R F M E T W A D M M P H H P H W V N E H L T I L Q F I A P S N P R 554
C A C A T T G G G A T G T C A A T G G G G A G G G T T A T G G A A A C G T G G G C A G A T G A T G C C T C A T C C C T C A T T G G G T A A C G A C A T T A C A T C A G T T T A T A C G G C G A C A C C C G C 73200

L R F E L H N P A F D F F V A P G O V D L P G P Q R P P E A M P T V N A T L R I I 594
T C T A G G T T G A T T A A C C C C G C T T G A T T T T T G T G C A C C G G G G A C G T A G A C C T T C C G G A C C G A C G G T C C C C G G A A C C C A T G C C A A C C G T T A A C C C A A C T T A C G G A T T A T T G 73320

N G N I P V P V P L C P I S F R D C R G T Q L G L G R H T N T P T A T I K A Y K O T F 634
C A A C C G G A A A C T T C C G G T G C C T C T A T G T C C C A T T C C G G A A G C T G T C C G G A A C C C A T C C G G T T G G G A A G C A T A C T G A C C C C G C A A C C A T T A A G C C G T A A A G G A T A C A T T 73440

E D R A Y P T I F Y M L E A V I H G N E R N F C A L L R L L T Q C I R G Y W E Q 674
T G A A G A C C G G C G C A T A C C C A A C T A T T T T C A T C G T C A G G G C T G T A T T C A T G G A A A G A A A C T T C T G C G C T T A C G G C A C T G T G A C A C G T G A T C G C G G G T A T T G G G A G C A 73560

S H R V A F V N N F H M L N H Y I T T Y L G N G E L P E Y C I N I Y R D L L Q H V 714
A T C C C A C A G G G G G C C T T G T G T A A T A A C T T C A T G T T A A T G C A T A C T A C A T T C T G G A A C C G G T G A C C G G T G A T T C G G G A T T A T A T A T A C G G G A T T A T C G C A G C A T G T 73680

R A L R O T I T D F T I Q G E G H N G E T S E A L N N I L T O D D T F I A P I L W 754
A A G A C T T A C G C C A A C T A T A C C G A T T T A C A T A C A G G A G G G G C C A T A C C G G G A G C C T C G G A A C C G T A A A T A C A T C C T A C C G G A T G A C C G T T A T T G C A C C T A T C A T G 73800

D C D A L I Y R D E A A R D R L P A I R V S G R N G Y Q A L H F V D W H A G H N F 794
G G A T T G T G A T G C G T T A A T A C C G G T A T G A G C C G C C G G A G A C C G A C T C C C C G C A T T C G T G T A A C G C G G G C A A A C C G G T A C C A C C C T C A C T T G T G G A T A T G C C G G C A T A A C T T 73920

Q R R D H N V L I H G R P V R G D T G Q G I P I T P H H D R E W G I L S K I Y Y 834
C C A A C G C C G C A T A T G T G T A A T C C C G G G A G A C C G G T C C G G G G A C A C G G G T C A C C C C T A C C C A C C A T G C C G G T G A T T G G G G A T T C T C T C C A A G T T A C T A C T A 74040

I V I P A F S R G S C C T M G V R Y D R L Y P A L Q A V I V P E I P A D E E A P 874
T A T T G T C A T T C C T G C T T C C C G G G G T C T G T G A T A C G G C G G T T A T G C C C T A T A C C C T G C G G C A A T T C C G G C T G A T G A A G A C C C C C 74160

T T P E D P R R H P L H A H O L V P N S L N V V F H N H A H L T V D G D A L L T L Q 914
A A T C A C C C C B A A G A T C C A A G C A C C C C T T C A C C G C A C C A C T G C T C C G A C T T C T T A C G T T A C T T C C A T A T G C A C A C C T A C C C T G A T G C G T G A T G C A C A C T A C A 74280

E L M G D M A E R T T A I L V S S A P D A G A A T A T T R N M R I Y D G A L Y H 954
A G A G T T A T G G G G A T A T G G C T G A C G A C G C G G C C A T T T G A T C A T C A G G C G C C C C G A T G C G G G A G C C G C C A C C A C C A G A A A T A T G C A G A A T A T A C G C G G A C C C T T A C C A 74400

G L I M M A Y Q A Y D E T I A T G T F F Y P V P V N P L F A C P E H L A S L R G 994
T G G C C T T A T T A T G G C A T A T C A G C C G C A T G C A T G A A C C A T T C A G C A C C G G T A C T T T T T A T C C C G T T C C G G T C A C C C T C T G T G C A T G T C C G G A A C A T T T G C C A T C A T T G C G T G G 74520

M T N A R R Y L A K N V P P I P P F L G A N H H A T I R Q P V A Y H V T H S K S 1034
A A T G C A A T A G C T A G G C G G G T T T G C C A A A A T G G T A C C A C C A T C C C T C T T C T G G G A G C C A A C C A C C G C A A C T A T A C G C C A A C C C G T G C C T A C C A T G T A A C C C A T A G T C 74640

D F N T L T Y S L L G G Y F K F T P I S L T H O L R T G F H P G I A F T V V R Q 1074
G G A T T T A A T A C T C T C T C T C T G G G G G G T A T T T A G G T T A C C A A T A T C T C T C A C A C T C A A T C A G C G G G A T T T C A C C C C G G G A T T T G C C T T A C C G G T A G T G C G C C A 74760

O R F A T E Q O L L Y A E R A S E S Y F V G Q I O V H H H D A I G G G V N F T L T Q 1114
G G A T C G C T T G C C A C A G G A C A C T T T A T A T G C C G A C G C G T G C T C T G C A T G A T T G C C G A C A A T C C A A G T A C A C C A T C A G T G C T A T T G G G G G G T A A C A C T T C C C T A A C C C A 74880

P R A H V D L G V G Y T A V C A T A A L R C P L T D M G N T A Q N L F F S R G G 1154
A C C C A G A G C T C A C G T G G A C C T G G G A G C T G G G A T A C A G C T G T A T G T G C C A C A G C G C A C C C T C T G C G A T T G G G G A T T G C C T T A C C G G T A G T G C G C C A 75000

V P M L H D N V T E S L R R I T A S G G R L N P T E P L P I F G G L R P A T S A 1194
A G T G C C A A T G T C A T G T A C A G C T A C C G G A T T C G T G C G T G T A A C A G C A T C G G G G G G T C C G T T A A T C C C A C C G G A A C C C C T A C C A T C T C G C G G G A C T G C T C T G C A T C A C C G C G C 75120

G I A R G Q A S V C E F V A M P V S T D L Q Y F R T A C N P R G R A S G M L Y M 1234
 AGGAATTGACGGGGCAAGCCTCTGTGTGAGTTGGCCATGCCGCTGCACTGACCTACAATTTAGAACTGCATGCAACTCTAGAGGTCGACATCTGGATGTTATAT 75240
 G D R D A D I E A I M F D H T Q S D V A Y T D R A T L N P M A S Q K H S Y G D R 1274
 GGGTACCGTGACGCCACATAGGGCTATAATGTTGATCACACACAATCGGATGTTGCTTACAGATCGACACTCTAACCCATGGCATCACAAAACATTACCGTGACAG 75360
 L Y N G T Y N L T G A S P I Y S P C F K F F T P A E V N T N C N T L O R L L M E 1314
 CCTATACACGGACATACAAACCTACAGGCCCTCTCTACAGCCCAGCTTAAGTTTACACCCAGGGAGTTAACACTAATGTAATACACTGGATGCCCTCTAATGGA 75480
 A K A V A S Q S S T D T E Y Q F K R P P G S T E N T Q D P C G L F Q E A Y P P L 1354
 GCGAAAGGCTGTCGCGCAAGCTCACCAGACTGAATATCAATTAAACGCCCTCCGGTTACCGGAAATGACACAGGATCCGTTGGCTTTTCAAGAAGCATATCCACACT 75600
 C S S D A A W L R T A H A G E T G A D E V H L A Q Y L I R D A S P L R G C L P L 1394
 ATGCTCAAGCGATGCCCATGTAGACGGCTACGGGGAGAACCGGGGAGATGAAGTCACTTACGCCAATATCTGATTCGAGACGGCTGCCCTTAGGGATGTCCTCT 75720
 P R - * TCCGCATAATTACACGCCACATACCCACTCCAATAAAAGCCCTGAGAGCCATTGGCATCTTACTGAGATTGGATACGCTGCCGACTTGGCTGTTCACGCTCCCTA 75840
 41 H A M P F E I E V L L P G E L S P A E T S A L O K C E G K I I T F S T L R H 38
 AACAAACATGGCTATGCCATTGAGATAGGGATTGTTACAGGAAATATCCCGGGGAAACATCTGCACTACAGAAATGTGAGGGAAAAATTATTACCTCTCACCGCTCAT 75960
 R A S L V D I A L S S Y Y I N G A P P D T L S L E A Y R M R F A A V I T R V I 78
 CGAGCTTCACTGGGATATGGCTGCTGCTCATATTACATTACCGGCTCCACAGACAGCTCGCTGTTAGGGCATACCGGATTCGATTCCGGCATTTCAACACGGTCATC 76080
 P G K L L A H A I G V G T P T P G L F I Q N T S P V D L C N G D Y I C L L P P V 118
 CCGGAAAGTTGGCGCATGCCATTGGCTGGTACTCCACCCGGGTTGTTATTCAAAATACATCCCCGGTGTATGGCTTGTAAATGGGATTACATCTGCTTACTCCGGTT 76200
 F G S A D S I R L D S V G L E I V F P L T I P Q T L M R E I I A K V V A R A V E 158
 TCCGGTCCGCAAGCTCAATCGCTGGACTCTGTAGGACTGGAAATTGTTCCCTTAACCATCCCCAGCCTTAATCGGAGAAATCATGCCAAAGTGGTGCACGGGCGTTGAG 76320
 R T A A G A Q I L P H E V L R G A D V I C Y N G R R R Y E L E T N L Q H R D G S D 198
 CGCACGGCCGGGCTCAAATTACCCACGAGTTCTACAGGGGGATGCTCGCTTATGGCCCTGATTCACACTGGAAACAAATTACACGGGATCGGATCGGATCGGAT 76440
 A A I R T L V L N L M F S I N E G C L L L L A L I P T L L V Q G A H D G Y V N L 238
 GCGCTATTCCGCACTATGGTTAAATCTAATGTTTCCATAAACAGGGGATGCTCGCTTATGGCCCTGATTCACACTTGTGTCAGGACACGCGTTATGTAATTTA 76560
 L I Q T A N C V R E T G Q L I N I P P M P R I Q O G H R R F P I Y E T I S S W I 278
 TTGATACAAACGGCAATTGGCTAGAGAACCGCCAGTTAAATAATACCGCAATGCCGGATTCAAGACGGCATCGCGATTCCCATATATGAACATCTGATA 76680
 S T S S R L G D T L G T R A I L R V C V F D G P S T V H P G D R T A V I Q - 316
 TCAACATCATCTAGCTGGGGGATACCTGGGACTCCGCAATTTCACGGCTGTTGTTGATGGACCCCTACTGTGTCATCCGGAGACGCCGACGGCTGATCAAGTGTAAACA 76800
 GGTGTTAATAAAACACACCACTGCTAGTTACATTTCACGGCTGTTGTTTATTAATAGGCATAAACACGGGATACCGGTATACATGACTGCCAATACACGGACATAATTATGCA 76920
 42 * - K I P M F V S D P I C S S G I Y V S H I L A 374
 ACCATCAGATCATCTGACATTGTTCCGGTGTACCTTACCCGTGAAAGTTTGTGCTAGATAACCCATACCGCTTAAATACCTCTGTCAGGTTACCAACTGTTACATGAC 77040
 Y M L D D S M T G T T G K G T Y T K T D L N G W G G G K I Y E T L N D L Q K C L Y 334
 TCCACGGGGTCTACACCTAATCTGTAGGATAACAGCTCTGTGAGCTATTATTCGGGAGTTAAATGTTAAACAAATAGCTACGGCCGGGTTTTGTTGTTGATA 77160
 E V P D V G L K V T L S V L E Q S A I T N G S N F R K V F Y D V A P T K Q K Q L 294
 AAAAAGGGTACGCCACGCTACATCCGGAGGTATGGAAATGATAAAACAGTAAACACTGGAGGAGATAGCAGTTCCTTCTGAGGAGACGCAAACGTTGCTATAGCCAC 77280
 L F F P Y A V S C G P P I S H Y F L L V P A S S L V N G K E L V A F Q Q A I A L 254
 GATATGGCAACTGCAAGATCTGGCTGTTCCCTCTATAGAAACGTGACGGTTGAAATGTTGGGGTGTAAAGCAGTATGGGCTAAGCATTGAGTAACGCAACGCCCTAC 77400
 S I A V A S D Q S S N G E I S V H Y N T F T N P H L A L I H G L C Q T V C R G I 214
 TCACTGGAGACGCTGCAAGTAAAGCTCAAGAAAAAGCTCCAATCAAATATAACCGACTTAAACGCCAACATCGTACACCGTACAGACGCTGTTGAT 77520
 E S S S T G T L A R L F F H E L G F I T W D S K Y R G V I A V G T G S A R T N S 174
 GTAAATGCGAGGTCTACGTAACAGTACACACTGACGATAATAGCACAATTGCAAGGGTACGGCCGATATAAAACCTCTACGGGGAGTTTTGTTGATAATAATGGCCGGTCA 77640
 T F A P D V V V Y L V S S L I A C N A V T S P R Y L I F R E R A T K T F L P R D 134
 AACCCACACCCCCAGAATTGTTACGCCACCTACAATTCTGCAAGGGAGTGGCCATAAATAATCTGCAAGTGCGCCATGGCTCATCTTGTGATGAAAACCGGCTTA 77760
 F G Y G G S N Q K R G G V I E Q Y F S D A M F L D A T R R M A G D M T I F V P K 94
 TTAAATACATAACGAAAGCTGACATGGCTATGTCATAACACGGGCACTGATGTCGCTACATATGAAACAGTTAACACTGATCCGACGATCCACGTAAGTTATAC 77880
 N L V Y C S C A T Y D S H A L V R P M H D D C V Y T V V N L L Q D S S G R L H Y 54
 AAAAAACTGACTCTGCTTCCGGTTGGTGTGAAACAAAAAAATTTTACAATTGGTTGATTTAAACATGGGACTATAGTTGATCACGGCATCGGTAATAATAGCTCA 78000
 L F S T S A K G T N T S S V F I I K C N T Q N L F G V I T Q V A D P R I F N A E 14
 TCCACAAAGAAGATAAAATCTGACCTCGGATACCCCTGGAAACGATGAAAGATATAGTTACCCACCAAACTGTTAAATGATCCCTAAATACACGTAACGTTGAA 78120
 D Y F L L N F D Q G R I G <---- spliced from 45

M E A H L A N E T K H A L W H N D H T K G L L H	24
TACGTACATTTCTTTTCCAGTACAACCATATCCGTGATAATGGAAGCCATTGGCAATGAAACCAAACATGCATTGGCATAATGTCACACAAAAGGATTACTACA	78240
V V I P N A G L I A A G I D P A L L I L K K P G Q R F K V E Y O T R Y H A T G Q	64
CGTTGATACCTAACCGGGGCTATTGGCCGGAAATAGATCCCGATTACTGATTTAAAGAACCCGGACAAACGCTAACGGTGAAAGTACAACAGAATATCATGCTACAGGTCA	78360
C E P M C Q V F A A Y I P D N A L T N L L I P K T E P F V S H V F S A T H N S G	104
ATGCGAACCGGGTCAAGTTGCCGCTACATCCGATAACGCCCTAACAAATCTCTAACACCAAAACCGAACCTTGGTACACGGTTCGGCACCGTAAATTCCG	78480
G L I L S L P V Y L S P G L F F D A F N V V A I R I N T G N R K H R D I C I N Y	144
GGGATTGATTATCATGCTGTTATCTTACGGCTGTTACGGGTTATCTTGTACGGTACGGATAACGAAATACTGGAAACCGGCAACGGTGTACGGTATTTGATTGTA	78600
A E L I P N G T R Y F A O G Q R V L L L C K Q L I A Y I R C T P R L A S S I K I	184
TGCGAACACTAACCCAAACGGAACCGCTTATTTGCTGATGGACAACGGGTACTTTATGCAACAGCTGATGGCTATCCGATGCAACCCCTGCTTCGTCATGTCATAAA	78720
Y A E H M V A A M G E S H T S N G D N I G P V S S I I D L D R Q L T S G G I D D	224
ATACCGAGAGACATATGGTGGCAGGCATGGTGAATCACACAGCTAACATGGGACAATATTGGACCCGTTACCCATAATCGATCTTGTACGACAGTTAAC	78840
S P A E T R I Q E N N R D V L E L I K R A V N I V N S R H P V R P S S S R V A S	264
CTCCCTGTGAAACACGCTACAGGAAATAATCGGGACGCTCTGCTAACATAAACGGGCGTAACATTGTTAACCTCCGACCCCCGTCGACCTCTAGTCCCGGTTGATC	78960
G L L Q S A K G H G A Q T S N T D P I N N G S F D G V L E P P G Q G R F T G K K	304
TGGGTGCTTAAAGTGCGAACAGGGCAGCGAACACTTCAACACAGTCGATCAAAACGGTCTTGTACGGCCTGACAGGCTTACGGGCTTACGGGAAAGAA	79080
N N S S A S I P P L Q D V L L F T P A S T E P Q S L M E W F D I C Y A Q L V S G	344
AAACAACTCGTCCGGCAGCATCCCACCTTACAAGAGCTTATGTTACCCGACTTGGCAGAACCCAAAGCTTATGGAAATGGTTCGACATCTGTTATGCCAATTAGTTAGCGG	79200
D T P A D F W K R R P L S I V P R H Y A E S P S P L I V V S Y N G S S A W G G R	384
GGACACTCCAGATTTCTGGAAACGGCGCCCTATCAATTGTCACCGCAGATTACCGCAGAACCTCCAGCTGGTGTATTGTAGTACTTACACGGATCTCTGCTGGGAGGAC	79320
I T G S P I L Y H S A Q A I I D A A C I N A R V D N P O S L H V T A R Q E L V A	424
TATTCACCGGAAAGTCATTTATATCAGCTGCAACAGCTATTATGTCGCTGGTGTATAATGCCGGTTGACAATCCCCAACGCTACATGTCACGCTCGCCAAGAGCTAGTCG	79440
R L P F L A N V L N N Q T P L P A F K P G A E M F L N Q V F K Q A C V T S L T Q	464
CGCTTACCGGTTTGGCTAACCGCTTAAATAATCAAACCCCTTACCGGCTTAAACCGGCCGAATGTTAAACCAGGTTTAAACAGGTTTAAACAGGTTGTGACATCGCTAACCC	79560
G L I T E L Q T H P T L Q Q L H E Y D I A D S Q T V I D E I V A R T P D L I Q	504
AGGCTTAAACGGAGTACAACGACCCGACTCACAAACAATCTGGAATATGAGATTCTCCCAAACGGTTATGTGAAATTGTAGCCGGCACACAGGACTTCA	79680
T I V S V L T E M S M D A F Y H S S L N M Y A V L A Y L S S V Y T R P Q G G G Y I	544
GACTATAGTTGGTGTAAACGGAAATGCAATGGATGGTGTATAACAGCTCTGTATGGCTTGGTGTCTGTCATCTGTATACACGCCAACAGGTTGGGGGTATAT	79800
P Y L H A S F P C W L G N R S I Y L F D Y Y N S G G E I L K L S K V P V P V A L	584
ACCCACTCCACCCCTTCCCATGCTGGTAGGTAACTGCTTATATATTTTGACTTATAATTCAGGAGGGAAATACTTAAACGGTCTCCGGTACGGCTT	79920
E K V G I G N S T Q L R G K F I R S A D I V D I G I C S K Y L P G Q C Y A Y I C	624
AGAAAAGGGTTGTTATGGTAACTCCACACAATGAGGGGAAATTATACCGCAGCCGGATTTGTGATTTGGAAATTGTTCTAGTATTACCCGGCTAACGTTACCCGTACATTG	80040
L G F N Q Q L Q S I L V L P G G F A A C F C I T D T L Q Q A A L P A S L I G P I L	664
TCTAGGATTAAACCGCAATTACAATCCATTAGTTACCGGGGGATTGGCGATTTGTGATTACCGATAACCTCACAGGCAACTACCTGATCTGTTAACGGACCTATTCT	80160
D R F C F S I P N P H K - *	676
AGACAGATTCTGCTCTATCCCAACCCCTATAATAAAATTAGTCIACATAAAAACATAACCCAGAATCTCTCATATGTAATTTCAGTCATTCTCCGGTTCCACCCCTCT	80280
T A A A T A A A A T A A C C G G T G G G G C A T T A A C C C A C A G T A C C C G G C G C A A T C C G C T A G A C T G T T T C G C T A T G G A A T T C A C C G C T A T G C T A C C G G C T A C G G G	14
T A A A T A A A A T A A C C G G T G G G G C A T T A A C C C A C A G T A C C C G G C G C A A T C C G C T A G A C T G T T T C G C T A T G G A A T T C A C C G C T A T G C T A C C C G C T A C G G G	80400
M E L Q R I F P L Y T A T G	54
TCCAGCGCCAAATTAAACCCCGAGGCAGTTCAGAGACTCTGGATGCTTAACCGCTGATGGATTATGGAACTCATCTGGATCCCCGGGTGAAAATAATGGCATACTGC	80520
F I T L R I A P F I P L O T D T T N I A V V V A T I Y I T R P R Q M N L P P K T	94
TTTATAACTTAAAGGATCGCTCCGTTATCCCTCTCAACGGATACTACTAATATGGCTTGTGTTACGGCACAATTACATCACCGGCCACGTCAGATGAACTTACCTCCGAAGAC	80640
F H V I V V N F H Y E V S Y A M T A T L R I Y P V E N I D H V F G A T F K N P I A	134
TTTCATGTAATTGTAATTGAGGTTACCGAGGYCTCGTACGCAATGCGCAGCTTAAAGAATTCTGGGTGAAACATGACCTGTTGGAGCACTGGAACTTAAAGAACCCGATCCG	80760
Y P L P T S I P D P R A D P T P A D L T P T P N L S H N Y L O P P R L P K N P Y A	174
GTACCCCTTCCCAACATCTTCCGGATCTGAGCAGATCCACCCCGCAGACTTACACAAACGCCAACCTAACGCAACTACTACACCCCGGCCCTCCGAAAATCCATACCG	80880
C K V I S P G V W W S D E R R R L Y V L A M E P N L I G L C P A G W H A R I L G	214
ATGTAAGTTATTCTCCGGAGTGTGGTGTACAGCAACGAAGGGCTTATATGTACTGGCTATGGAACTTAAAGGCTATGTCATCCGGCAGTGGCAAGGTTAGAACCCGATCCG	81000
S V L N R L L S H A D G C D E C N H R V H V G A L Y A L P H V T N H A E G C V C	254
CTCTGTTAAATCGACTCTCGTACGGCAACGGATGTCAGTAACTAGGTTACCGTGGGGACTGTGCTACGTCGGTACCTGTCACATGCGGAAGGGTGTGTG	81120

L H 24
TACA 78240

G Q 64
ATCA 78360

S G 104
CAGG 78480

M Y 144
ATGA 78600

K I 184
AAAT 78720

D D 224
ATGA 78840

A S 264
ACATC 78960

K K 304
AGAA 79080

S G 344
GGCG 79200

G R 384
GAGC 79320

V A 424
ATCC 79440

T Q 464
CCCC 79560

I Q 504
ATCA 79680

Y I 544
ATAT 79800

A L 584
CCTT 79920

I C 624
TTTG 80040

I L 664
TTCT 80160

676
CTCT 80280

T G 14
GGGG 80400

T A 54
CTGC 80520

K T 94
AGAC 80640

I A 134
TCGC 80760

Y A 174
ACGC 80880

L G 214
TTGG 81000

V C 254
TGTC 81120

W A P C M W R K A G Q R E L K V E V D I G A T O V L F V D V T T C I R I T S K 294
T T G G G C T C C G T G T A T G T G G A G A A G G C C G G T C A G C G G G A A T T A A A G T G G A G G T A G A C T T G G C C C A C G C A G G T C T T T G T A G A T G T C A C C A C T G C A T T G C A T T A C G A G T A C T A A 81240

N P R I T A N L G D V I A G T H A S G L S V P V N S S G W Q L Y M F G E T L S R 334
A A A T C C T C G C A T T A C C G C A A T C T G G C G A C G T T A T G C G G G A C C A C G C C A G T G G T C T C T G T A C C A G T A A A T T C A T G G G T G G C A G C G T T T A T G T T G G A G A A C C T A A G C C G 81360

A I I N G C G L L Q R I C F P E T O R L S G E P E P T T T - 363
G C G T A T T A A C C G C T G T G T C T G C G G A A T T G C T C C C G G A C A C A A G T T A C C G G T G A A C C G G A C C T A A C A C C A C T A G T A T A C C T T A C T C A A C C G C T T G T G G A A 81480

45 AGGTATATGTCACATTACAGTAATAATTAAAGGTTAAATTATAAACACTCACGTTTGTTGACTTGACCCGAACACCGCTGTGCTGTAGAGACCCGTCGGTAATGAAAACCGT 81600
spliced to 42 ---> N T N H S S A F V A T S Y S G D T F S F T 332

A A T A G A T T C G C C T T T A C T G A T C C A C G T A A T T G C C C C A A C C A C T G T G C C G G A A G C A T G T A T C C C T C A A A C C G G T C C C T G T G C T A C G C G G T A C C A C T T T 81720

I S E G K V H D V Y N A G F W Q E L R S K I G E F V P E T A K R I H A T Y G V K 292

A A T C C T C A A A C G T G G C C A T T A C T A A G C G T A T T A A G G T A C A A G A A C C A T G T T T C C C A T G T C T A C G C G G T A C C C C A A A C A C G T T G A G T T T G T T G A A G T G T C T A A A C A C T G T C 81840

I G R F T A M V L A I L P V L F M T K G H R R P V L F V T S K Q K F H E L V S D 252

A G A A A C A C T T G G C G T G T T A A A C T G T A C G C G A A A G C A G T C A A C T C T G T G C C G C A T G A C C G C A T G A A A T A A A T G C G T G G T G C A T G A G G A T C A T T T T G A A C A G 81960

S Y S P T N F V T R L F C D V R D A H D G L L V Y S I F H T T H M L I M K O F L 212

T T C C A A C G T C C C C T T A T A T C G C C A T A G A T T G G A A C G T C A A C C T T G C C G T T T G C C A T G A C T T C C A C A C T C T C A A A C A G T G T T C C C A A C A G G T A C G A A A C C G T T G T G T 82080

E L T G K Y R G Y I P Y D V K A R K G H S G C E E I S E F S T E V L Y S F R Q T 172

A A A G G T A C A C T G A C G A A A C A T C C G A C G A G A A A C C G C G G A A T G Y G T T C A T A A C C C G C T A T A C C G A T T T C G A T G A G G T G C T G C T T T C C G G T G A A T T C A A A C T G T A C 82200

F T S L Q C F S D S L S F A R F T H M V G S Y A N R H P A A E E P S Y E Y F Q V 132

A C T A C T G A C G C C T T T T T A T C A C C G G T T A C G T T G C A T T A C C G G A A T T C A C T A C T G G G G G T A C A G T T G A C C C T G T G A C G T A G A A A C C G C G C C A A C A T 82320

S S V A K K L E P S V N A N Y S Y R W P K L V V N P T C N Y G T S S L F A G F M 92

T G C C C G T C G A C G A T G A C C G G A A C A G T G G A A T A T T C A C A A C A G T G T G A A C C G C T A T C C C T A C G C G C T T A C G C G C T T A C T T A C G C A A A T T C A G A A A C G G 82440

A R R A T A S F L P I Y E C C N H L T G I G P I V A Q O H R R T Y D I A F N L F P 52

G A T T T G G G T T G C G T T C C C A G A G A C C C T G C C G G T G G A C A C C G G G T A G G G G A C T C C A C G C G T A C T C C C T A C G C G C T T A C T G C T T A C A G C T T A A A T C T C A G A T T C T C 82560

I Q T A N G L S G Q R T S C P T P S E L T G F R E D R R R K L R E F Y R V S E E 12

A C C A G C G T A C G C C A A C A T T A C A T G A C A T T A A C A T C A T C A C G G A A T C C G C C T A C T C T G T A A G C A G T A A A C A G G A A G C C G C G T A C T T A C G T A C T C G T A C G T A T A T 82680

1 G L T R G F M I L S W

46 M S G H T P T Y A S H R R H R Y K L V E A H N R A G L 27
C A T A A C A T T T C A G G C C G C A T T C A C T T G G T C A T G T C A G G G C A C T C C A A C T C A C G C T T C T C A T A G G C G T A A C C G T C A A C T A G T T G A G G C G C T A A C C G C C G G G G T T A T 82800

F K E R T L D L I R G G A S Y Q D P A F V Y A F T A A K E A C A D L H N Q L R S 67
T T A A G A C C G G A C C T G C A T T A T C C G T G G G G T G C G A G T G T A C A G A T C C A C G C A T T T G T G T A T G C C T T A C T G T G C A A A G G G C T G C G C G A T T T A A C C A C G C T C G C T G 82920

A A R I A S V E Q K I R D I Q S K V E E Q T S I Q Q I L N T N R R Y I A P D F I 107
C A G E T C G C A T A G C T T C A G T G A C A G A A G T C G T G A T A C A T C A A G G T T G A G G A C A A C A G T A T T C A A C A G C G C T A T A C G C A C C G G A T T T A T T C A C G C G G A T T T A T T C 83040

R G L D K T E D D N T D N I D R L E D A V G P H I E H E N H T W F G E D D E A L 147
G C G G T T G G A A A A C A G A A G C G A T A A C C G G A T A A T A G A C A G T G G A A G C G C G T G A G G A C C G A A C T C G A C A C G G A A A T C A T T C T G T T G G A A G A G C G A C G G T A C 83160

47 M D A D D T P P N L Q I S P T A G P L R S H H N T D G H E P H A T A A D Q Q 38
L T Q W N L T T H P P T T S K Y L Q L Q D L C V P T T I P T D M N Q M Q P O P I S 187
T T C A C A A T G G A T C G T G A C G A C A C C C C C C A C C T C C A A A T C T C C A A C T G C A G G A C C T T G C T T C C C A C C A A T C C G A C G G A C A T G A C C A A A T G C A C C G C A G C C G A T C A G C A 83280

E R E S T N P T H G C V N H P W A N P S T A T C M E S P E R S Q Q T S L F L L K 78
K H E N P P T P H T D V - 199
A G A C G A G A C A T C C A C C A C C C C A C A C C G G A T G T G T A A A T C A T C C A T G G C C A T C C G T C A C T G C A A C A T G C A T G G A A T C A C C A G A C G T C A C A C A G C T T A T T T A A A 83400

H G L T R D P I H Q R E R V D Y F P Q F N K P P N V F R I S K L S R L I V P I F 118
G C A C G G T T A C G G A G A T C C A A C A T C A C C G G A A A G G G T G G A C G T T T C C A C A T T A C A A C A C C C C A T G G G T T T T A G A A T T C C A A T T A C C C G T T T A T T G T A C C C A T T C T C 83520

T L N E O L C F S K L Q I R D R P R F A G R G T Y G R Y H I Y P S S K I A V K T 158
C A C G C T A A T G A C A G T T A T G T T T C T A A T T A C A G A T T C G A G A T A G A C C C A G G T T G C G G G A C C G T A T G G G C G T G T C A T A T A C C C A T C G T C A A A A T G C T G T A A A A C 83640

Y A 174
ACGC 80880

48 M D S R V F N R E L I N A I L A S E G S I R A G E R L G I S S I V C L L G F S L 198
C A T G G A C A G T C G T G T T T A A T G A G G G T T A A T C C G C A T T T A C G C A G G G G A A G G C T A G G T T T C T A C G A G C A G G G G A A A G G C T A G G T T T C C T T A G G T T T C G T T 83760

L G 214
TTGG 81000

V C 254
TGTC 81120

O L A Q A L T F L N R T C G L T H L D V K C G N I F L N V D N F A S L E I T T A 277
A G A T T T G G C T A A C G C G T T G C G T T T A A T C G A C G T G C G G C C T G A C C C A C T G A T G T G A A T G T G C C A T A T T T C T A C G T C A C A T G A C A G A G A A A T T G C C C A T G T A T T T 84000

VZV DNA sequence

1791

CGCTGGAGATTGAAACCGATATGGTCATAATACTGGTCCATATACTGGAGTGGCGGGGTCCCAACGAGTAGCACAGGCCACGAGCCTTCCCAGGGACGGCTCTATTAAACCGTGTATAA
 RPSQVSIPDYLRLDMYSHAAATGLTACAVLTGLSPGILVHIY 87000
 295
 ATGCCGAAAATTAACTCTGATACTATAAGATAACAAGTACTAAATGAGCATGGCCACGGACCGATGACCAGTCCCGTATGAGTATGTTGCCTGCAACAGTC
 HALTLESVILYVVSVSLLTSFTSHAVSVSSWLGTHLHNAYLE 87120
 255
 CAGCTTAATGATACAAATAGGATACATATGCCATCAACGGCCATCAAATCAGAACACTGGCGCGTAGGCCCGCAAGGGCATATAAAAGACGCTCTGCTGCTGATAATTGCG
 LNLSVFLICIAMLAANLNVFVARTPGALRYLFVSQQRNLNA 87240
 215
 GACCGCTTATGTTGCTTGTGCTCCAAATTTCGGCTGCCAAAAAAATGCTGTAATATTACACTTGTCGCAAAATGCTCAAGATAATGTCAGGCCACGCCATTGCTGTAAGC
 YAKINTEDLKGRGFYTTFIVSTAFHGLYLTAAVGYIQLKYA 87360
 175
 TAATAATAACACACGGCTTAAATACCAACATGACAAAAGCCCCAAAAAAAGTGTGTTGGGATCTACAACTAACCTGCAACACGGGAGCTTGGCCGACAGTGTATTTCGTTTC
 LLLVVANLLWLSLLGWFLTTPDVVLWAVGSSQRVRQNKT 87480
 135
 TCGGTGTYATAACGGGGCGTGATCAGTGTATACCCGCAATTGGCCATTGGCGTTAAAGCCTGTTAAAGCCTGTTAGTAAGTAAATGCCAACACGCTATGTTGCTCCAAAA
 RHIIAATILTYVANAMATLATYTYTFAVVSHPELFLVPASY 87600
 95
 TCCACCTCTATTCGGCCACATACCCCCCATAGGGTGGCTTAATACTCATACCTACGGCAGCATAAAAACAGGGATCCGGTATATTCAAGAGGGCCGTTAACGT 87720
 GGRHGSWVGGDLTANFLEYDGVVAAYFCPIGTYESSAAILT
 55
 ASCLAGGGACATTACCGACCCAAAGTGAACATCATCACCTGAATTATCCTGCAATTAAAGCTGATTCATTGATGATCTAGCCTCCACCTCGGGTGTCTGTTGCTGACGG
 ALLNVAGLTFMWVQIWIIRNHLTDMDMQHDLAEVEPPTTDYP 87840
 15
 M SP N T G E S N A A V Y A S S T O L A R A L Y G G D 27
 CGAGACTTTTCAGAACCGCGCCCCCTCTTITGAGTCTCCCATGTCCTCCAAACACGGGGAGGCAACGCCGCGCTATGCGTCTCAGTACACAGTCGGGGGGCTGTAACG
 SYKESRPGKKQQTGM 87960
 1
 LVSNWIKHTHPGISLELQLODVPVKLICKPGMSOTRPVTVVRA 67
 TCTGGTTCTGGATTAAACACACCCACCCGGAAATTGGCTGGAACCTGCAATTGGATGTTCTGAGTAAACTAATAAAACCTGTTATGTCACAAACTGCCGGTAACCGTCTGACGTC
 88080
 PMGSGKTTTALLEWLOHALKADISVLVVSCRRSFTQTLIQR 107
 CCCTATGGCTCTGGATAAAACACACCCCTGGTTCAGTACACGGCTAAAGCAGATTACCGCTACTGGTTGCTCATGEGCCGTAACCTTACCCAGCCTGATTCACCG
 88200
 FNDAGLSGFVTYLTSETYINGFKRLIVQLESLSRHRYSSSEAI 147
 GTTAAACGATGAGGGCTCTGGTTCAGTACATATTGAGCATCCGAGACATATTATGGGTTTAAACGCTTGGATTGTCACACTGAAAGCCTACACGGCTATCCAGCGAAGCTAT
 88320
 DSYDVLILDEVMSVIGOLYSPTNRRLSAVOSLLYRLLNRC 187
 CGCAGCTACGAGTAAATACTGGATGAGGAATGTCACTGTTGGACATTACTCCCTACAGCTGCTTCCGGTGTGATGCTTACCTGCTTATATCGCTCTTAACTCGCT
 88440
 SOIIMANDATVNSQFIDLISGLRGDENEIHTIVCTYAGVGF 227
 TTCTCAATTATCCTGGGATGCTACAGTAAACTCGAGTTATGATTAACTCCGGATTTGGCAGATGAAACACATACACAAATTGTTGTCACATCGCGGGAGTTGGGCTC
 88560
 GRTCTILRDNGIDTLLVRYIKRSPEHEVDVRTIHLQLRGTF 267
 CGGAAACTGCGATCTGGCTGATATGGGATCTGACAGCTGGCTGCTGAGTCACTTAAACGATCTCTGCAACACGGAGTGTACGTAACACACTACGTTGGACATTTC
 88680
 EALALRLOCGHNICIFSSSTLSFSELVAOFCAIFTDOSLILN 307
 CGGACTAGCACTACGATTACAATGTTGGGATACATCTGTTATTCATCAACTTTATGGGCTGGAGCTAGTGTCTGCAATTTCAGACTCTATTC
 88800
 STRPLCNVNEWKHFRVLYTTTVVTVGLSFDMAHFHSNFAY 347
 CTCAACTCGCCCTATGTAATGTAACGAACTTTCGCGTGTGGTCACTACCGCTGTCAGCTGGATGTTGACATGGCTCATTTCTACAGTGTGGCT 88920
 ICPMSVGPDPMSVYVQSLGRVRLLLLNEVLYVVDGSRTRCG 387
 CATGGACCAATGTCATATGGGCCGATATGGTATCGCTACAGCTTGGCTGCTGAGGAGATGCTCTTGGAGATGCTGCTGATGCTGCTGCTCAAGGACAGATGCGG
 89040
 PLFSPMHLLNFTIANKFQXWFPPTHQITNKLCCAFRQRCANA 427
 ACCCTGTCGCCAAATGTTACTAACTTACATGCCAAATAATTCACTGGTTCTCACACACCCAAATAACTAACAAACTGCTGTCATTAGGCAACGATGTCAAATG
 89160
 FTRSNTHLFSRFKYKHLFERCSLWMSLADSINILQTLLASN 467
 ATTACACGCTGACACCCATCTCTCAAGTTAAACAAACCTTGGAGAGATGCTCTTGGAGATGCTGCTGCTGATGCTTACAAACTCTTGGCTCTAA
 89280
 QI LYVLDGMGPITDVSPPQFCAFIHDLRHSANAYASCWRS 507
 CGGAAATTGTTGATGTCATGGGCTCAATACGGAGCTTCCCGAGTCATTGGCTTACAGCTCTGACATAGCGCTAACCGGCTAGTGTGCTGCTGCTGCT
 89400
 LRQONDSCLTDFGPSPGFMADNITAFMEXKYLNESINTEEQI 547
 TCTTAGCAGGAAATGACAGCTGCTGAGCTGTTGGCCCTTCCGGATTATGGCCGATAACATTACCGGTTTATGGAAAAGTATCTTGGAGTCATTAACCGAAGAACAAAT
 89520
 KVFKAACPIEQPRLYNTAAILGACIRIPEALEAFDVFQKI 587
 TAAGTATTAAGCCCTGCTGCAATGACAGCTAGACTGCAATACGCCAATATTGGGGGCTGTTACGAAATCTGAGCTGTTGGAGCTTGGAGCTGTTCCGCTGTC
 89640
 VTHYASGWFPVLDOKTGEFSIATITAPNLTTHWELFRC 627
 ATACACCCACTACGCTCCGGTTGGTCTGGACAACCCGGGAATTAGCATGGCAGACTAACTACCGCCCAAAATTAAACACACATGGGAGCTGTTCCGCTGTC
 89760
 VIAKTLKWMPSSTECECVTOVLDLTDINTLNFNOHGDSLALQIF 667
 CTATATGCCAAACACTCAAGGTTAATGGCTCCGGCTGCTGTAACACAGTTTGGATGAGCATGGACATTAACTACCTGCAACAGGGGATTCGCTGCTGCT
 89880

E Y W R C N V T D A K I I L H R P V W R T T G F L D G C H N Q C F R P I P T K H 707
TGAGGTATGCGCTGTAACGTTACTGACGCTAAAGATTATAAAACCGCCGGTTGGCGAACACCCGGATCTTAGATGGATGCCATAATCAATGCTCCGTCCAATCCCTACAAAACA 90000
T A T A C C
Y T
E Y N I A L F R L I V E Q L F G A R Y T K S T O T F P G S T R V K N L K K K O L 747
CGAATATAACATGCCCTATTCCTTTAATTGGGAACAAATTATTCGGCCGGCTAACTAAAGTACCCAGACCTTCCGGGAAGTACTCGTGTGAAAACCTAAAAAGATCT 90120
C T T G C /
Q V
E T L L O S I N V D R S A C R T Y R Q L Y N L L M S Q R H S F S Q O R Y K I T A 787
AGAAACTTTACTTGATCAATTAAACGGATCGTCTCGATTCGGCTACCTCCGGAGTGTGATAACCTGCTTAGGCCAGCCATTCGCTCTCAACAGEGTTACAAAATTAAGTGC 90240
A C T T A /
K V
P A W A R H Y Y F Q A H Q M H L A P H A E A N L Q L A L S E L S P G S W P R I N 827
CCCCGCTTGGCACCGCTGTTTCAGCAGCATCAAATGCACTTGGCCCGATGCCGAAGCCATGCTACAAATTAGCCCTATGCCAACGTGCCGGGATCGTGGCCGGATAAA 90360
C C C A A M
L F
G A V N F E S L - 835
CGGGGGCGGTAATTGGAAAGTTATAACCCGTTAACCATATATGGACATCCATAGGGGGGGTACATAAAACTAACGCTCTGTACACACAAAGGGCTCTAACATGCACTGAAC 90480
A A T C T /
D I
N D A T Q I T L V R E S G H I C A A S I Y T S W T Q S G Q L T Q N G L S 36
CACAACCAAGCTATGGACGCCAACCGAGATTACCTGGTTAGAGAACCGGACACATTGGCCGCAAGCATATAACACATCTGGACACAGCTGGAACAAATTAAACACAGACGGCTTCC 90500
A C G C T T /
A I
V L Y Y L L C K N S C G K Y V P K F A E I T V Q Q E D L C R Y S R H G G S V S A 76
GTGTTATACTACTTATTATGCAAAACACTGTTGGGAATACGCTCTAGTGGCAATTACCGTACAAAGGAGTTATGTCGCTACTCCAGGATGGGGGAGTGTCTCGC 90720
A G T T Y /
N V
A T F A S I C R A A S S A A L D A M P L E P L G N A D T W R C L H G T A L A T L 116
GCAACGTTTGGCTATCTGCAGGGCGCGCTCCGCTGGCTTAGACGCCCTGGCCCTTGAACACTGGGTAACGCCAGACCTGGCTGGCTGGCTGGCACTGGCCACTTTA 90840
G T G T T /
T I
R R V L G F K S F Y S P V T F E T D T N T G L L L K T I P D E H A L H N D N T P 156
CGCCGGTATTAGGGTTAAATGCTTTATTCGCCAGTAACTGGAGACTGATAGAATACAGGTTCTGTTAAACAAATCCCGATGACACCCGTTGAATAATGCAACACCCCA 90960
T C C G C /
R I
S T G V L R A N F P V A I D V S A V S A C N A H T Q G T S L A Y A R L T A L K S 196
TCTACCGGGTATTGGGGCTAATTCCCGTGGCCATTGATGTTAGCAGCTAGCCGATGTAACGCCAACAGGAAGCTGGCTAGCCGCTAGCCGCTAGCCGCTAGCCGACTAAATCT 91080
C T T T C /
R E
N G D T Q Q Q T P L D O V E V I T P K A Y I R R K Y K S T F S P P I E R E G O T S 236
AACGGTGACCCAGAACAAACACCTTAGCCTGGAGGTTAACACAAAGGCTACATCGCTGGAAATATAAGGCTAGCTTTCCCCCTATAGAGCGGGAAAGGCCAACCTCC 91200
T A G G A /
Y S
D L F N L E E R R L V L S G N R A I V V R V L L P C Y F D C L T T D S T V T S S 276
GATTGTTAACTTGAGAACCCGCGCTGGTTCTAGGGCAATCGCCATTGTTAGGTTACTCTTACCGTGTGTTAACTGTTAAACACGGATTCACCGTTACATCTCC 91320
G C C C /
P A
L S I L A T Y R L W Y A A A F G K P G V V R P I F A Y L G P E L N P K G E D R D 316
CTTCAATATTAGCAACATATAGCTGTGGTACCCGGCGGTTGGAAACCCGGGTTGTCGCTCCAACTTTCGGTATTAGGCTGGAAACTCAATCCGAAGGGTGAAGACAGAGAC 91440
A A A C I /
L R
Y F C T V G F P G W T T L R T Q T P A V E S I R T A T E M Y N E T D G L W P V T 356
TACTTTTGACTGTGGATTCCCGGATGGACACTCTCGGACACAAACTCCAGGCTGCAATCTTACCGCAGCTACGGAGATGTACATGGAAACGGATGGGGTGGCCAGTAACC 91560
G T O T O /
R T
G I Q A F H Y L A P W G Q H P P L P P R V Q D L I G Q I P Q D T G H A D A T V N 396
GGTATTACGGCCCTTCATTATCTACCCCGGTTGGGACAGCATCCCGCTTACCTCCGGGCTGGAGCTATTGGGAAATCCCTCAAGATCTGGACATGCAACTGTCAAT 91680
A C A T /
V D
W D A G R I S T V F K Q P V Q L Q D R M M A K F D F S A F F P T I Y C A M F P M 436
TGGGACCGCCGGGGATCTACCGCTCTAACACGGCTGAACTACAAGATCGTGGATGGCAAGTTGATTCACGGCTTTTCCACGATATACTGGCTATGTCCTGG 91800
G C C C /
A G
H F R L G K I V L A R M R R G M G C L K P A L V S F F G G L R H I L P S I Y K A 476
CATTTAGGCAAAATCGCTGGCTAGAATCGCTGGAGGAATGGGGTGCTAAACCCGGTTGGCTGGCTTTGGGGGTTACGGCACACTCCGAGTATAACAGCT 91920
A T A T /
I D
I I F I A N E I S L C V E Q T A L E Q G F A I C T Y I K D G F W G I F T D L H T 516
ATTATTTTATGCAAAATTAGCCTTGGCTGAACAAACGGCTTGGAAACAGGGCTTGCTATGACTTATAAAAGGATGGTTGGGGATCTTACCGGATATACAGCT 92040
T G A G /
H S
R N V C S D Q A R C S A L H L A A T C E R A V T G L L R I Q L G L N F T P A N E 556
CCCAATGTATGTCAGGCACTGGCTGGTTGGGCTTAAATTAGGGGACCTCTGGCAAGAGCAGCTACGGGCTTACCGGATATACTGGCTTAACTTACCCGCCATGGAA 92160
C C G C I /
R O
P V L R Y E G V Y T H A F T W C T T G S W L W N L Q T N T P P D L V G Y P W R S 596
CCGGTACTCCGGCTGGGGTGTGATCACTCACGGCTTACCTGGGCTGGCTGGGATGGCTGGGAAATTACAAACACGGCTCCGGATTAGTGGCTGGCATGGCAAGT 92280
C A G G C G C G C A G A T T A A G G A C C G C T T C A G G A C T C C T A T G T A C C G C A A C A A A T T C G A G A A C C G A T A C G G A A A T T C G G A A C A A A G C A T A C C G 92400
T T A T /
N Y
Q V V E A A R K T Y V D F F E H Y F D R R Y T P V V W S L Q E Q N S E T K A I P 676
CAAGTTGTGGGGCTGGCAAGAAAACATACTGCTGATTTTGACATGTTGATGCGCTTACTCCGGTATCTGGGATCTGGAGCTTCAAGGAGAAAATCCGAAACAAAGCAATACCG 92520
G A C /
S I
A S Y L T Y G H M Q D K D Y K P R O I I M V R H P N P H G P P T V V Y N E L L P 716
GCATCTTATGACATACGGCACATGCAAGATAAGGTTATAACCAAGACAGATAATTGGTCGTAATCCAAACCCACATGGACCTCTACTGGTTACTGGGAATGTCACCA 92640
G C A /
A I
S C A C I P P I D C A A H L K P L I H T F V T I I N H L L D A H N D F S S P S L 756
TCGTTGCGCTGATTCCTCCATGACTGCGCTGCTCATCTCAAGCCCTTACACAGCTGGTCACTATTAAACCATCTCTAGATGCTCATATGATTTCAGTCATTC 92760
C G T /
R V
K F T D D P L A S Y N F L F L - 771
AAATTACTGACGATCCCTTGCTCATATAACTCTTGTGTTTATGACAACCCAGCCGAAACACCCATCTTAAATAAAAGGTTATTACTTACACCCGTGGTAATTTC 92880
A A C C /
L I
53

TATACTTCAAAATACTGAACTTTCCGGTGTACCATGGTGGGATTTAACCAAAAAATAACGCTTCTGATATTCCGAATCTGTAAGGTCCATTAAACAATCCGGGGGTA 93000
 YTEFLQVNKPPTVNTRLWWFVYVSKQYESDRLPGPNLLGPPLY 284
 CTTGCACACACCATCTGGACAGGGGGGGGGTCCTGGGGCAGGTCAAACGCTGACCCACCCACATGAATAATAGCCTTATAATATTGGGGCGTCCAGGCTGAGGGTCAAGTA 93120
 QVVGDPCPPTGHPLDFRQGVGCSYIAKIIINPATGPQPPL 244
 ACTTAACAAACATATAATGGCGCAATACGGGGTTTGTAAGGGGTGTTAACAACGACATACATTAGAGTGTAAACACCATAAAACTCCCTCATATAAAAACCAGCGATT 93240
 KVFMYHPLVTRTKTFPNNODVYVHLTNLLWLVGEYLFRRMKE 204
 CCAGGGTCTATTGACACTCAACGGCTAAAGATAACAGACAACTGACAAACAGGATGGAGATGCCCGGAGGGCCAATGCCCTTCAGGATACATTAAAATAACACATAAGGTA 93360
 LPGIQCEVRRLYVSLQVFLSPSAGSP6IGELYNLIVCLTF 164
 AATCTAGGACATTATCCGGCGGAATAGAGTCATCGGATAGTTAACAGGCGGGAGGCACCCACCGTACACCCCTATCTTCACCGCAGTTAACGGAAAAAAATAATCCGGGA 93480
 DLVNDPDRFLTMRYILLRPPVGVTVYRDEVATLVSFIFGRF 124
 ACCTGGTGGAGTAACACACTCATGAGTAACGATCACAGGACACCTCACTGAAATCACCATTCAACACTAAACGGCTCTGGTTCGGTTAACGGCAGTGATAACACAG 93600
 APQTVCENYRDCSVESSDGNLVLVLTQEHEPKVRLSVVS 84
 AGTTGCCAAAAGCGTGGCTTCAACCGGTTACCTCCGGCCCTCGCATAAGAATCTGGTATTGCTGTTCTAAGATCTGCTGACGTCGCTCACATCCAACCCCTTCGCTC 93720
 HALFRPKLGTVERAECSVFRPIAQIRLDEIVDSDVLDGEEAR 44
 -SRRDRRECGVGGRRS 756
 54
 GTGTTAGTAAGTTGCGATCGTACGGTCAACCTAAATGCTGGTATATTTCGGACATCCCACATGGGATCTCCCGCCTCGGTTGCTGAAATTTCAGTAAGGTGAA 93840
 TLLNDITVSCGLISPINIGSMGDAANGAGGTQEFKIWYPRI 4
 TNTLQRDNRQLRFHQTYKNRVDWRGDGRRNARIKDLTS 716
 TCCGCTGATTTACCTTGTGACCGTAACTCTCAGGGGGGTGCTCTTCAAAATGGGATAGGTTTATATCCAACTGATGTTGGTTTTGGGTTCCGGGATT 93960
 ROM 1
 DAANVKHYRLREPPPHGKMFHSLNKYGVHMYQNNNIKNPEPI 676
 CTTCGTCATCTCTGTTAGGGTCAGGCAACCCCAAGGAAGACTGGTCTCCGGCCCGGTTTATTACCTCTGCGGAACTGCTTACATATAATTCGGATTGGGATAAA 94080
 REDDETDPDPLGWSPSPTRRPGRKIVEARVQMEYLIRIQL 636
 TAGGACTCTGTTCTGCCCTTTAAAGACCTGGCATAACTCTCTGACCTATGACCTCTGAGTTACCAAGAATCTAATCGGGTGGCCGTAATGAAATGAAAAATAC 94200
 YSETRAKKFIAQCLEEESRHVESQTVLFGLRTARLIFSFY 596
 GGGCAACTAGTAAGGATTGACCGATTTGAAATGACAGAAATTCTGGCTTGTATTGTTACCCGGTGAAGCTTAAACGCGAACAGTTCTGTTCCATAGCTGAC 94320
 PAYLLSISANSYSVSIEQGQONNNVRHLKFCRVLEQKWL 556
 AAACGTTTATATCATCTCCATAAGGGGGATAAACAGGATTGAAAACATTGGCAATATGCACTACCCCTTATATGCCGTAAGATCTAACCTCTGATTTAACGGCAATA 94440
 LRKIDDGYPPIYRSQFSNAIYADDGIIIGTLTDIVEHNLD 516
 CGTGTCTCTGCCATTGTAATATGACCTTTAGATGGCTTATTTTACCCCTCTCCCGTAACCGGTTGACTCTGGGCTTCGGTAGATGCTGTC 94560
 RTEEAINTHSKSPKIKVREERLRKLEGEEKFQLRETLD 476
 ACATCCTGAGACCCCTCAATGGTTGAATAATTACATACCCCTCGAGCATGGCTTGTACTGTTACCCGGAGTTTAAACGCACTTGAACGTTGTTCCGGACATT 94680
 VOKLGEITKFLNNVYGELMNGNISHNVYVSTKFASQVNTTGSM 436
 GCCCCCGGTTAAAGGGTGGCTTGCACCCGGTTGATGTCACCGATCCACTCTCTCCAGAATGTTGCGCTTCTAGATAAGGAACGTAACGGTTCCGTA 94800
 AGGNFSQNAKGFGPQSTDVSGSGELIHNGTEELYSRVYET 396
 ATATCTCAACATCTCATGTTTAACTATTGCTTACAAGCTGACCCGGCGATCCACGGCGTGTGTTGATCTGCTTCTGCCATTATACGATCAACGGCACGCTGCTG 94920
 IDGVHRMNKLNVILKVLRSAASGARTTDNEGMIRDVARTS 356
 TGAGATCTATCATCTCATTCCGGCAGCTTAAACACGGCAAAGGGGCTGTTAAACTGGCAGACGGCAGCTGTTACGTAATGCTAACAGGCAACCTCCCAAGAAC 95040
 HSRDDENRRGILVRLPATNLYOCVRAHERLAYCALVEGSL 316
 CGCTGTAAGGGTGGTCAAAACTACACCCCTCCCAACATACACGGGGGACACGACCAACACTCTCCCTCATGCCGTTACATCATCTTGTCCATAATTACCTGGTTAA 95160
 RQLPSDFVVGEGCVVPPWVLCLEGKMGTVDDKAMILRRNY 276
 TTATAAAAGACCGCTTCTCATATACTCAACATTTGCTACACTCAACAGGCTGACACCCGGCTCCCTCTGCCAACGGTGTACGTCACGTTACATCTGG 95280
 NYYLRLTRDYDMIAVNOMCEVLSVVAAGRALTADAVKL 236
 GACAGTCTGCCGCTTGACCCATATACTGTTAATGGTCAAGGGGTTCCATTCTGTTGATCGTACCTTCTACACGGGCAACATACGCTACAGGCTATCCGACGTTTG 95400
 SLEAAQGMYTNLPAPTGHNQESRVKRYVVPVIGVCAINDVYK 196
 GCAAAACGGACCCCTCCATTAAACACTGGTATAGAGAACCGGTTTCCACCGAGAAACTCAAGTAACGATGACTGTTGACGGCTAGGGTCAAAACCTGATGTGCAAGC 95520
 AFGVGRGNLGLGSTYLCGTTIGRLFELLSSQLTQRWTTEFVQHAL 156
 CGTACGGCTTCTGATTCTCCACATACGGCTAACGGCTAGAGCCGGCATGCAAGGTTACATTGTTGATGTTGTTCCAACTGCTGCTAGGCTCTACCCGAGTTGCA 95640
 RYAESESEGCLGYREALAGAHLNCCQOIHEDALDEYRTAD 116
 AACGGCTTCATCAAAACGGTTGCTGAACAGTTCCGTAGACCGCTATATGCCCTGCTACGGTATATCCAAAGTCACGGCTAGGGATTTCGAAAC 95760
 LANMLVTAQVQRIVTESTRVASYIGQGDTYGFODGALIKRF 76

AACATCACTTGGCGGTTGGGTGATTAAACATCCAGCCTATCTCCGGAAATGTCACAAAACCTATATCCGGGGCGTACTCATTCAGTATATCGAACATGTTCTGTATTGGTC	95880
LMSQTTPHILNWGDDEEPFTCFGIDPAYENWYIDFMMNKYQD	36
TTGGGTACTTCATTCAAGCCCTGGTCAATAGAAAAGAACTTGCTATCCTTTCTTCACTACCGGAACGTATTAAAGAGAGCTTATTCCGGCATTGAAAACACGATGAA	M K 2
NPNSGNLGODISVSSAIRKEESGSSNNFLSTIEAM	96000
1	
R S I S V D S S S P K H V F N P E T P N G F D D S V Y L N F T S M H S I Q P I L	42
AGATCAATTCTGTAGACGTTCTCACCCAAAACGTTTAAATCCAGAGCCTGAACTGGATTGATGACAGTGTATTTAAACTCACCTCATGCATGCACTTCAACCTATCC	96120
S R I R E L A A I T I P K E R V P R L C W F K Q O L L E Q P P E M Q R N E L P	82
CTCACGGATTGAGAACCTGCCGAAATTACGATTCACAAAGAACGCTGTCGGGTTGTTAAACAGTTACTCGAAGCTGCAAGGCCCTCGAAATGCAAGGAGATGAGCTCC	96240
F S V Y L I S G N A G S G K S T C T C I O T L N E A I D C I I T G S T R V A A Q N V	122
CTTCCTCGTTTATTAAATTAGCGGAAATGCCGCTCCGGAAAAAGCAGTGTATCCAACGCTTAACGAAGCTATCGATGATTACCGGATCACCGGGTGTGCGCCAAATGT	96360
H A K L S T A Y A S R P I N T I F H E F G F R G R N H I Q A Q L G R Y A Y N N T T	152
TCATGCTAAGTATTCAACGGCTTATGCCGAGTCCTGATAAACACAATCTTCATGAAATTGGTTTCGCGGAAATCACATTAGCTGCTGGGGTTACCGCATATAACTGGACTAC	96480
T P P S I E D L Q K R D I V V Y W E Y L I D I T K R V F Q M G D D G R G G T S T	202
GACCCCCCTCTATTAGGAGGACTCGAAAAAGAGATATTGTATACACTGGGAAGTTTAAATGATAACAAAACGAGTGTTCATGGGAGACGGTCCGAGGAAACATCGAC	96600
F K T L W A I E R L L N K P T G S M S G T A F I A C G S L P A F T R S N V I V	242
ATTTAAAACCTGCGGCAATTGAGCTTGCCTAATAAACCTACAGGCTCAATGCGGAAACGGGTTATCGCATGCGGTTCCCTCCGCTTTTACCGGAGACGTTATTGTTAT	96720
D E A G L L G R H I L T A V V Y C X W W L L N A I Y Q S P Q Y I N G R K P V I V C	282
TGATGAGACGAGATTGCTAGGGCGCTATATTCTCACCGGGCTGTTACTGTGTTGGCTTGTAACTATATATAACAGGCTTACAGTACATACCGGCTGAAACCGGTACATGATG	96840
V G S P T Q T D S L E S H F Q H D M Q R S H V T P S E N I L T Y I I C N Q T L R	322
CGTCGGTTGCCACCCAAAATCTGAGCTGTTAACATCTTCAACATGACATCGAGCTGACACGGTTCACACGTAACCTCTAGTGAATAACTACGATATAACTCGCATAAACTCTGCG	96960
Q Y T H I S H N N A I F I N N K R C O E D D F G N L L K T L E Y G L P I T E A H	362
TCAATATACTAACATCTCACATAACTGGCAACTTATTATAACAAACGATGTCAGAGGAGCATTTGGAAATCTTTAAACCCCTTAGAGTACGGGCTACCTATTACCGAAGCACA	97080
A R L Y D T F V V P A S Y I N N H P A H L P G W T R T L R Y S S H K E V S A Y M S K L	402
TGCGCGCTGTCGATCATTGTTACCTGCATCTTAAACATCCTGCTATTAACATCCTGCTATCTCCGGATGGACGCGCTGTATTCGTCGCTATAAGGAGGTGAGCCGCTATAGTAAGT	97200
H A H L K L S K N D H F S V F A L P T Y T F I R L T A F D E Y R K L T G Q P G L	442
ACACGGCATTAAAACATCGAAATGACCATTCTGTTGCTTACCGACTTATACTCATCCGCTAACGGCATTTGATGAACCGCAATTACCGGACAACCCGACT	97320
S V E H W I R A N S G R L H N Y S Q S R D H D M G T V K Y E T H S N R D L I V A	482
TTCTGTGTAACATTGGATACGGGAAACTCCGGCTGTTGCAAACTTCCCAAGCGGAGATCATGACATGGAAACAGTTAAACGAAACACATCAAATCGCGACTTAACTGAGC	97440
R T D I T Y V L N S L V V V T T R L R K L V I G F S G T F Q S F A K Y L R D D S	522
CCGTACAGACATCATCTAGCTGCTAAATAGCTCGTAGTTGTAACCCAAAGACTACGCTAGTTAGTATTGGATTAGCTGCTACATTCAATCGTTGCAAGGTTTACGTGACGACTC	97560
F V K A R G E T S I E Y A Y R F L S N L I F G G L I N F Y N F L L H K N L H P D	562
CTTGTGAGGCTCGAGGAGAGACATCCATCGAAATGCTTACCGGTTCTGTCAAACCTTACCTGGGCTGCTGTTAACTTAACTTTGTTAAATAAAACCTACCTCCGA	97680
K V S L A Y K R L A A L T L E L L S G T N K A P L H E A A V N G A G A G I D C D	602
TAAGGTATCGCTAGCATACAAACGGTAGCTGCTTACCCCTGAGTTATGCTGGAAACAAACAAAGCCCCCTAACCGAGCAGCGGTATAGGGGGGGTGGGGATTGACTGTGA	97800
G A A T S A D K A F C F T K A P E S K V T A S I P E D P D D O V I F T A L N D E V	642
TGGTCAGCTACTCTGGCGATAAACGCTCTGCTTACCAAAAGCCCCCGAGTCACAGGCTCCATACCGGAAGACCCGGATGATGTAATTTCACCGACTAACGACGAGGT	97920
I D L V Y C Q Y E F S Y P K S S N E V H A O F L L W K A I Y D G R Y A I L A E L	682
TATTGACTTGGTACTGCGAGTACGAAATTCTCTATCCAAATCATCCGCTATGAGGTCTGCTGTTAATGAAAGCTTACCGATGCTGCTATGCGCATATTAGCAGACGCT	98040
F E S S F T T A P F S A Y V D N V N F N G S E L L I G N Y R G G L L S L A L Q T	722
TTTCGAAAGCCTTACACCCCGCCCTTACGGCGTATGTCGATAATGTTAAACCGGAAGCGCCTTGTGATCGCAATGTCGCGGGGGGGCTGTTATCTTGGCATACAGAAC	98160
D T Y T L L G Y T F A P V P V F V E E L T R K K L Y R E T T E M H Y A L H V P L	762
AGATACGCTACCTTGGGTTACTTGTGACCCCTGAGGACTGACCCGAAAGACTGTCAGCTGTTCTGTTAATGAAAGCTTACCGATGCTGCTATGCGCATATTAGCAGCT	98280
M V L Q D Q H G F V S I V N A N Y C E F T E S I E D A E L A N A T T V D Y G L S	802
TATGGCTTACAGGATCAACATGGTTTGTCATCGTAAACGCTACGTTGTAATTACCGACTATAGAGGATCGAGAATTGCGCAATGGCCACACCGGTGACTATGGCTTAG	98400
S K L A M T I A R S Q G L S L E K Y A I C F T A D K L R L N S Y Y V A M S R T Y	842
TTCTAAACTAGCCATGACAATTCCACCGTACAGGGCTGAGTTAGAGGAGCTATCTGTTTACCGGCTAAACTCGCCCTAACATGTTGATGTCGCGTACCGT	98520
M K N P Q K L A I T F L P L Y V I P T Y T L C I	24
S S R F L K M N L N P L R E R Y E K S A E I S D H I L A A L R D P N V H V Y Y -	881
CTCTCTAGGTCTTAAATGAATCTAAACCCCTACGGGAAACGATGAAAATCCGCAAGAAATAGCGATCACATTCTGGCGCTCTAGCTGATCCCAACGTCACGTTGTTATT	98640

K A L Y K N T H A G L L F S F L G F V L N T P A M S I S G P T T F I L Y R L H 64
 AAGCATTTGTAACCCACGGCATGGGGCTTGTCTCATTTCTAGGTTTGCTTAAATACACCCGCCATGAGCATCTCGGACCCCCAACGACGTTTATTTATAGGTTACATG 98760
 G V R R V L H W T L P D H E Q T L Y A F T G G S R S M A V K T D A R C D T N S G 104
 GGGTAGCGGGGTTCTTCACTGGACTTACCGGATCATGACAACACACTACGGATTACGGGTGGGTCAGATCATGGGTGAAGACGGGACTGATGATACAATGAGCGGTG 98880
 G M I V L Q H T H T V T L L T I D C S T D F S S Y A F T H R D F H L Q D K P H A 144
 GTATGATCGCCTCAACACACCCATACTAGTGACCCCTGCTAACCATAGCTGTTACTCATACGGATTACGGGATTCTACAGGACAACCCACCGAA 99000
 T F A M P F M S W V G S D P T S Q L Y S N V G G V L S V I T E D D L S N C I S I 184
 CATTGCGATGCCCTTATGCTGGTCTGACCCACATCTAGCTGATACAGTAATGGGGGGGGTACTATCCGATAAACCGAAGATGACCTATCCATGTTATCTAAC 99120
 V I Y G L R V N R P D D Q T T P T P H Q Y T S Q R R Q P E T N C P S S P O P 224
 TTATACCGGTTACGGGTAACAGCTGACGATCACGACACACCAACCCGCACCGATACATCGCAAGGGCGCAGCGTAAACCAACTGTCCTTCAAC 99240
 A F F T S D D D V L S L I L R D A A N A - 244
 CCTTTTCACATCAGACGACGACGCTTCCTGTTAAATACGGGACGGCAACACGGTAAGACAGATTCAAGACTAACATTATCCAACCTGATTACATTCTACCGGAATAAAC 99360
 ACACAAAAAAATTATTTAACGGCTTTAATTGAAGACACCTATCTCTTAACTCGGTGATGAGCCTTGCAGGTTGGGTGCGCGCTTACCGGTTATACATAACCGATTACCGTGT 99480
 57 - R Q H A K C T P H R A E G T N Y M V S K G H 50
 K R C D S W K G T Y T Q L V V N N H S L I L E R V F K Q H E G P T A V S A N G 10
 TTACGGCAGTCYGACCATTTACCGTGTGTAATACGACGCTGTTGTCGGCAAAATTAACTCGGCTACAAATTCTGATGTTCCCCGGCGTGGCAACGCTGGCATTCGA 99600
 AACACATTACGGCTCTCGTACGCTCATGGCTAAAGTATTTCCTTATGTAAGGACACGATCTACACCGTAAACCGTAAACTCATACACAAACAC 99720
 F V H R E R V D M 1
 58 - T R T R G H G S N E T H I P Q D T L T N E K H L L V R D L A T F E Y Y F V 185
 TGGTACCAACGGACGGATTTCCGCTGGTGGCGGTTAATCGCGGAGGTCTTGCACGAATCTCGTACAGTAGGTTCTGACACGGGGTGCAATGGGTTTTGACACAA 99840
 P V L P R S K G D T S R T Y Y R R P R R A R I S E Y L L N R V R P A H T K Q C L 145
 CACAAACATTTCAGGCTCTTACTGGGATTGATGGGATTGAATTATTAGGGTCAGCTGGTTGGGTTGGCTACCGGACCTTCTGACACGGGGTGCAATGGGTTTTGACACAA 99960
 V F M Q L S K H S S P N F K N K S L T V H K Q R S V G R G S F A A T K T C A V T 105
 TATTTCACAGGGCTTACCAACGCTGGCGGATGGTTCGGTTAATGCTCCTCCGCAAGTTCGTCATAATGATGACCATGACAACGATCATGGTACATGGTCTTTGGT 100080
 I E C A N H V L S R R I T D T L Q R R G L E D I S S V M F L T D F P V Y D D K T 65
 TTTCTCAATACAGCCCGCTGCCAATCGAAATTTCATTTGCTACACGGCTATTTCTGTAACACTGTTGCTGGCTACCTGTTAAAATTGGGATCGAAC 100200
 K E I C G A H G I P F K E N A D V S N E T F D N Q Y S H O S A V Q K F N P I S C 25
 CGGTCCACGATGCAATCCCAACCCATTGAGGAATGCCGTCGGTACGGAAGGGCAACTCGAAAACATTATGGTACGCAAGGGTCGATTGGAGTGTATATAACACTCCAATCG 100320
 59 - I V S W D 301
 P G R H L G W G W Q L L A T P V S P P L E S F M 1
 ATCTCGGGTTGCCCTTACCGCTAAACTCATGGCTGAACGAAATGTCGACAATTGGAAACACGGGACAATGGCGACGGATGCCGTGTTAGCACCAGATGACATCTT 100440
 I E P E G K R T F Y E N A Q V F H R C N R F P V R S L P S P H A T L V L H C R 261
 GAATTCCGGTTGGGTCTCTGTCGATGCCACCCACAGCATAAAAACTAACCCCTGACGGTCTCGCATAACCTCTGAGCAGCGTTGACCCAGCCGCCAGCCTAAGTATACA 100560
 S N P Q T T K Q A H A G V L M F V L G T R N E C L R Q L V R Q V L R G W G L Y V 221
 TGGCAGCCCCGGAGTCCCGGACGAAACGGTAAGCGTGGTATTACGCAATAACACCCCTGCTTGCACCAACTCTCCAGGCACTCCGTAAGCTGGGGAGTCATATTGGGTATGATTCAGT 100680
 H S G P T G R R V T L T H L L V G Q R A W S E L C G H T P P T M N P Y S E M 181
 AGGGCCGAAAAATTAAAGACTAGACGGTGGTGTATGCCACGTTTACACTAACCGTAGCCCAGTGTGCTGTCGGGGTATGGGTATGGTACCCAATAATTACAACCGGA 100800
 L A A F I N K L S S P P T I G R K V S F A L G H A H G A T P Y P D Q G I I V V R 141
 ATGCTCTGGGGTCCGCAAATCGCTCCATGCAAAATACCGCTGAGTGGAGTTCTCCCTGAAATTAAAAGACGATTGTTCTAAACCTTCCGCTACGGCT 100920
 I S Q P G C F R T M A F I D G T S P L I E E G S N L L R N Y E L F I G K A Y P E 101
 TTAAGTCTGCGACAAACGGTCATACCAACTCAGGGAAATGTTAAACTTGTCTGAAACCTCAACCGAATCCAGTGGCAAGAGACGGGGGGGTAACGTTCCGCTGTAATGATGTGAC 101040
 K L E D S L L D Y W E P S I N F K S F V E V S D L Q S S V P T F T E T D Y H H S 61
 ATGTTTTACTGAAAGTTGGGGGTCTAGCTAACCCCAAAAGCGAGCCCGGGGTCGCTTGGGGTTTTGGTAACCGGATGGGCAAAACATAATGCTTTGACATCCGAT 101160
 H H H N L K F T P P D L V G L P L G R P R K R T K K T V P H A L V Y I D K S D S 21
 AGTTTCATTCATGGCATACGGCTGGAAACACGGCTGGCTCCCCAGACACATCCATTTCGGGATTTGGAGATGGGAGATGGAGTAGAGTCTACCCATACACGGAAAGGGCATCCAA 101280
 L K M E N A Y A N S C V T P E G S V D M 1
 60 - Q C V R Q F L R D A G V C G H E P I N T S S P T S D V W V G S L A D L 124
 CAAAGCATCGCTATGCTCCGCTTTATGTTCTACCAACAGATTGTGCGCAGCCCCCTTAAAGGTGACGTATGGATTGTCAGTACGCCATTGGTGTCTTAAACCAAAAGTAAAC 101400
 L A D R I D G S K H E E G V S Q A L G K L T V V P N T H Y A M Q K D K F W L I V 84
 TTCCGGTACTGGACATTGGCTTACCCAGATTCCCGATAGCCCTCGTGGGTTGATACCGGGGGGTCGCAAGTCCACGCCCTATACCGATGACACGGACGGTCCGTT 101520
 E P V P C K T K V V I G S L A E S L N S V P P A A Y D W A E Y V S S V C P E T I 44

IRL ----- IRS
 GAGGGCTGCCCTCCTCCGGGTCGCCGGGCCAGAAACGGGGGGGGTTATTCGGGGGGTCCGACCAGCCCCTCGCCGCCACAGACAGACACTTT 105120
 TTCATAAAAACGTTCCGCTTTATTAAACAACAAACAGTCCCGCGCCAGTGGCCCTCACGAGAAAAGGAGGGACTCCGTACCCCCGACTCTCGGGGGCTCCCTCCCCCGCCCT 105240
 - G R S Q P P S R G G R G 1299
 CCCACACATCGCTCGTCCGAGGAGCAGGAGAACAGCTCCACCTTGACCGCGGCCAACACCCACCCGGCTCGCAGCACACCCGGGCCACCGACACGATGTC 105360
 G W V D D E D E S S S S S L E V K V A P R L G V R R D R L V G P A V S V I S 1259
 ACCCCAAAGGATGACCCCGTGTCCCCTCGTCCCGCCCCCTCCCTCGTGTCCCACGGCTTCACACCCCACTCCCAATGTCAGCTCAAGCGTGTCTCTGTCGTGCG 105480
 V G F S S G P A D G D D G G G E E S D W A D E C G V E W O D L E L A H E R D A 1219
 GTGCCGCGCTGTGCCCGCCCTGGGTTCTGACGGCTTGACGCCAGGCGGCCAACACGGACGGTGTCCGAGCTCCCTCCACACCGCTCCGGGCCAACACGGGC 105600
 T R R Q R G A Q T E S P R E S G G H H G F V F R T G D S G E L V T E A A G F G P 1179
 GGCCACATTACTCTGGGAATCGGGGGGGGATTCGGAGGCTCGTCCGCGACGCATACAGCGCACCAGGACCGGACGGGTYGAAGCAGGAGCTGGTTCTGCGGAGGGTGGGT 105720
 P N M V R P I P P L A N R A E D A S A Y L A V S R G A V P P L V L P E A A P D P 1139
 TCCAGCAGGGCGTGGCGCAAACACCCCTGCCCAAGGTGGTAGCTGCCGCCCTCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 105840
 E L L A H R C F V R A N T P V D G A E P G A A G G P R R G E P L F V P R L A A G L 1099
 CCCCATCGGTTCTGCGCGGTGGCTATGTCGCCCTCGTCCACAAAGTCGGCTGCCGAGCCCAAGCCCCAGACTGTCGGCGAGGTCTTGCACACCGTCAAACCGGCAACCGCAG 105960
 G W R N A A R H S H A A E D V F D A A G L G L G R S Q R A L D K C G D F G P L V 1059
 TACTGCCGATTACCGGGGACAGGGGACGGGGCTGGGTTCTGGGCGCGCGGGTAGACACGGTGTATGCGACGTCTCCACCCGGGACACACAGGGTTGTGCGCCCGGGTACAGG 106080
 Y Q R Y E R P S L P V R T K P G A R T C V T Y A V N G G R C L C L P Q E G P Y L 1019
 TTGGCAAACCGAGTCTCGATAACGAGCAAACCTCGCTGGCCAAGGTGCGGAGATGCAAACACGGCCGGCGAGTCCTTCGACGGGAGTCTGGGATCGGAGCGACGGCCCTGG 106200
 H A F A T E I R A F S A P G F T R S S A F V A R A L G E T V A S D P W R V V A Q 979
 GCGTCGGTCGGCGGGGCCGGAGCTACGGTAGACTGAGACAAAGCGGGTCATCCCTGGGCCACCTCTCGAGGGGACCCGGTCAACACAGCAACCGGCGGGGAGGGCC 106320
 A D P R A P A R V V V H Y Q S L A P G D R P W R E L A V A D L V L L R R R A S A 939
 AACCGCGAGCCCTAGATACTCGACGGCCCGAAGGCCAGGTCTGGGTGACAGTAATAAACGCCCGGGGTCAAAGGGAGACGTCCGGGGGGGTCCAGTCCCGGGCCAG 106440
 L R S G L Y E V A G A F A L D R T S L L L V G R A N H A S V D P P P G T W N G A W 899
 GCATGAGTGTGGCAGGACAACCGGTTACTCAGGGCTGCCAGGACACAGACAGTCCCCCTGGGATGGACTCATGAGGCTCCGGATCTGCGAGGGTGTCTGAGGGGGCC 106560
 A H T S P L C L R N S L A A L V V S L G G R S P S W S P G P D T A L T S E L P G 859
 TTGATGTCCTCTCGGGGACACGGATCTGAGATGTCAGAGACCTACATCTCGGGGCTCTGGGATCTGCGCATCCAGGGCACTCTGGCTGAGCGCTGGGACAGGGGATG 106680
 H I D E G P L P D Y I I L L R V D E P D P I Q R M M A C R R R L A E Y G S P P G 819
 AACCGCTGGTCTCTCCGCCGGGAGCCGCGAGGGGATTTCCGCAAGGGCGGGGATCAAAGCTTAGCGCAGGGCGCAGGGCTGGGAAACAATGGTGTGCGACAGGGGATG 106800
 F R R D G G G P R R A A I E A L A G P D F S L A P R W A T P F L P D D V L R A I 779
 GTTCGGGGGATCACAGTACGGCTTGCGAGCTGGTGCAGGGGACGGGGTATGCGAGGGCCCCGGGGATAACGGCAAATCCCGTTGGGGCCGGTCCAGTGGCATCGTTATT 106920
 T E P T C Y A K R A Q D S P V P T H L A G R P I R R F G G N P A P G D L P M T I 739
 ACGCGGGGGGATCACACAGGGCGAGGGTATGGTGCAGGGCTCGGATAACCCGCTCTGGGCTTGGGAAACCATGATCGCTGCAACCCGGCGCCGACGGGTGCTCCCTA 107040
 V A P P D V V P G S T I T V P E S V R R K A K S V Y H D D A Y R A D A V P T E R 699
 ATCTTGCGAGGGCTCTGCTCTGACTGGCTGGGACTTGGCTGGCGAGTTGTAACAGATCATCGGTGGACACAGAGAAAGAGGCTGGCGGGGAGGGCTGAGGGTCG 107160
 I K D L L S R S E V P Q S K R K R P T R L R D O P P C V S L S R A A A S P Q P D 659
 GGAGCTGTGGCCGGGGTTGGTGGAGAAGGGTACGGCCGGGAGATCCGGCCGGGACTGGACCCGGTGGCTGGGATATGCCATGCTGGCAAGGGCTCGGGAGACTGTAGG 107280
 P A Q T A P T T P S P H G R S I R A A P S S G T A E P Y A M S A F A R R L S Q L 619
 ATAAAGTGTGTTGGGCGGGTGTGAGGAGGGTACGGCCGGGAGATCCGGCCGGGACTGGACCCGGTGGCTGGGATGGCCATGGCTGGGCTACATGGGAAACGGGAGAC 107400
 I F H K Q A R D Y R R S M A V A A A A H P L A N L A E G T A M A E G V H P V P S 579
 GCTACGCTCCCGTAAACGGCTAACCCCGCCGGTGGCAACAGCTTTGGTAGACTGGTCAAGGGCGAGGGTGGCTGGGATGGCCATGGCTGGCTACATGGGAAACGGGAGAC 107520
 A V S G T V A T G A R G P P L L K O Y F Q N L A S N Y G T L K P N Q L W A I P D 539
 CTGCTGGAGCTAGATCAGGTTAATCAGGGCGGGACTGTCAAGGGGATCTCCAACTCCGGCACGTAAGGGGACGGGTTGGCTGGGATGGCCATGGCTGGGCTACATGGG 107640
 R D P C Y I L N I L A R Y Q R A P D G L E P V Y L P V P E T S A E Y R A R A A R 499
 ACAGCCTCATCTCCAGTGGCTCTGGTCTCCGGGAGCTGGATGGAGGGGGTGGCGATCCGGGCAAGGGCTTCCGGTGGGCTACATGGGAAACGGGAGGGCT 107760
 V A E D E W H G E R T E G S P G F R V R N S P L P A S G P W P S G D P H W M L P G 459
 GACACGGGGGAAATTACGGGGTCTGGATCGGGCAGGGAAATGATTCTGCTGGGGCCGGGCTCCGGCAAGACGTTGGCTTCACTGGATCGGGATCGGGCT 107880
 S V P P I I P T R S R P L S F S K Q R Q R G P E G A L R K T K R I R P D P G S I 419
 GATCGATATCCGGTGGGATATTTGTTCTGCGACCCACCATCTTGAGTCGAATCATCCGAAATTGACGGGGAAAGGGGGTGTCTGCGCTGGGACCTGCTGCTGATGGT 108000
 S R Y G P Q I N Q K T S G G D N S D D S H S P S P A H E R G S R S G T T E S 379

N T P S P K R P Q R A I E R Y A G A E T A E Y T A A K A L T A L G E G G V D W K 260
 AACACCCCGTCCCCAAAGAGACCCCAAGCGGTGCCATCGAGCGATAACGGGGTCAGAAACCCCGGAATAACAGCCGAAAGCGCTACCGCGTGGCGAGGGGGGTAGATTGGAAG 111360
 R R R M E A P R R H D I P P P H G V - 278
 CGACGTGACACGAAGCCCGCGCCGATGATAACCGCCCCCATGGCGTAGTCTTATAATAATAACATGGTTGGCTGTCTTTTTGATGTCGTGTTGGGAGT 111480
 64 M N L C G S R G E H P G 12
 GGGGTGTTGGATATTAGGGTAGAGGGTGCTGGTTGAACGCTCCATTAAACCCACGGGTCCCCCACGGGCGTGTGGTAGATCTCGCGATCCCGGGTAGACCCCCGG 111600
 G E Y A G L Y C T R ' H D T P A H Q A L M N D A E R Y F A A A L C A I S T E A Y E 52
 GGTAATAATGCCGACTTACTGCACAGCACGATAACCCCGCAGCAGCTCTCATGAACGCGCAGCGTACTTCGCCCGCGCTAGGCCATTACCGAGGGCTACAGAG 111720
 A F I H S P S E R P C A S L W G R A K D A F G R M C G E L A A D R Q R P P S V P 92
 GCTTTATACAGCCCCCTCGAGAGACCTGCGCGAGTTGGGGAGGCCAAAGGACCGCTTCCGGAGGGCTGGAGACGGGATAGACAACGTCACCCCTCGGGTCCG 111840
 P I R R A Y L S L L R E Q C M P D P Q S H L E L S E R L I L M A Y W C C L G H A 132
 CGCATCCGAGAGCCGTTATCGTTATCGCAGCAATGCCATGCCGATCCAACTCGCATGGCTCAGCGAGCGCTGATATTGATGGCATATTGGCTGTGTTGGGACACGCC 111960
 G L P T I G L S P D N K C I R A E L Y D R P G G I C H R L F D A Y L G C G S L G 172
 GGGCTCCGACTATTGGATTGCCCCGATAATAATGCATCCGCCGAAATTATGACGCCGGGGAAATTGACAGGCTTTGACCGTACCTGGCTGCCGGTCCCTGGG 112080
 V P R T Y E R S - 180
 GTCCCAAGAACCTACGAGAGATECTGACACCCATCCCTTATATAGAAAAAAAAAAATAACACCGGATAAAAGCGTACTGTTTTTATTAATTAACACGCTCGGC 112200
 TTGCCCCGGTCCGGTATCAGGGTCTTATCATACACCGTGAACCTGAACACCCCGTGAATCCCTCCATCGGTTACCAACTCTCTCCGATACCGTAGATCCGAGCTCCG 112320
 IRS ----><----US
 AAATCGTCACCTATCCAAACAAATTGTGACGTTATATCCAAAGGCAAAGGCCGCTCCGTATAGCAAATAACAAAGACAATTATAGCTAATATAACAGAATTTCGATGATAT 12440
 65 - G V F Q S T I Y G L A F A A G T H A F V F V I I L T I Y C F K K R Y 68
 ATTATGTTGATATTCCAACTCGGCCAAAATTCTGCGCTTTCATCTGCTATCATAATAACACTTTAGCCGAAACGGCTCGGTGATGGCTGTATGCTGTATT 112560
 I K H Q Y K G I R R L F E D A T E N E S D S Y Y C K E A S R S P Q I A T I T N 28
 TTGGTTCGCTCGGGGGTTACACCGCTTCATCACTAGTAAGCCACGGCTCACCCTCATGGTGTGTTGCGCCATAGAAATCCAGATTGTAAGGGCAGCAGGCTAGTTAAAAGT 112680
 N P Q A R P T V V A E M L L A V A E G E M T H N Q G A M 1
 GTTAAATACCAACACCTTTGATATTATATACATGCAAGATTCTAGATTATTCATCAATAGGTGTTAAAGCGGTTTCATAAACGTTGTCAGCTACCGACATTCTCACAAAGAGG 112800
 TAAAGTTACCTTACGTTATTAAATAACATGAGACATTATAATAATCTAGGAACAATCAAATCTATTTGTAAGTTGTTAACCCCTCCCTTTGTCATATCTCCG 112920
 66 N 1
 CCTCTATAATCGGATCACTTATAAGTGTGCGGTGAGTATATTGTAAGTTGGACAAAGGTTTGGCTTACAACTATCAACATAAGTGGGTATACAAGTATAATGA 113040
 NO V O D A T D T F V G Q G K F R G A I S T S P S H I M Q T C G F I Q Q M F P V E 41
 ACGACGTTGATGCAACAGACCCCTGGACAAGGAAAGTCGGCGCCATCTCACACATCACGTCACATTTATGCAACAGATGTTCCAGTGAAA 113160
 M S P G I E S E D D P H Y D V N M O I Q S F N I F D G V H E T E A E A S V A L C 81
 TGTCGCCCCGATAGAATCTGAGGATGATCCAAATTGACGTTAACATGGGATACAGCTTTAATATTTGATGGTGTACAGAACTGAAGCCGAAAGCTGTTGGC 113280
 A E A R Y G I N K A G F V I L K T F T P G A E G F A F A C M D S K T C E H V V I 121
 CAGAGCAGCGTGGATTATAACGGGATTGTAATTAAAAACGTTAACACAGGGGGGAAGGTTGCGTTGCGGTGATGGACAGTAAACATGTAACATGTTGGCTATTA 113400
 K A G Q R Q G T A T E A T V L R A L T H P S V Y O L K G T F T Y N K M T C L I L 161
 AGCGGGTCACGTCAGGAACGCCACCGGGGTTAACGGCTAACCCACCATCGGTGACGCTTAAGGACGTTACGCTTAACAAATGACATGCTTATATTAC 113520
 P R Y R T D L Y C Y L A A K R N L P I C D I L A I Q R S V L R A L Q Y L H N N S 201
 CACGTTACCGAACAGATTATCTGCTATCTGCAAAACGCAACCTCCCATATGTCACATTACAGCGATCTGTTACCGCGGTTACAGTCTTACATAAAACAGTA 113640
 I I H R D I K S E N I F I N H P G D V C V G D F G A A C C F P V D I N A N R Y Y G 241
 TTATCAGCGTATAAAATCTGAAATAATTATTAACACCCAGGTGATGGTTGTTGGGGACTTTGGAGCAGCGTGTGTTCCCGGTTGATTTAATGCAACAGGTTGGCT 113760
 W A G T I A T N S P E L L A R D P Y G P A Y D I W S A G I Y L F E M A T G Q N S 281
 GGGCTGGAAACATGCCACAAACTCTCTGAGTTGGCTAGAGATCCATATGGCTGGACATATGGAGTGCGGGATGTTATTGAAATGGCTACAGGACAGAACCTG 113880
 L F E R D G L D G N C D S E R O I K L I I R R S G T H P N E F P I N P T S N L R 321
 TATTTGAGACGAGCTGGTTAGATGCCAAATTGAGCTGGACACTTAAACCTTATACGACGCTGGACTCATCCAAATGAAATTCCCATACATCAAATCTCGTC 114000
 R Q Y I G L A K R S S R K P G S R P L W T N L Y E L P I D O L E Y L I C K M L S F 361
 GACAATACATGGTTGGCAAACGGCTCTCTGCAAACCCGGATCCAGGGCATGTGGACAAATCTATGAGTGGCAATTGGATTGGGATTTGATAAGATGTTG 114120
 D A R H R P S A E V L L N H S V F Q T L P D P Y P P N P N E V G D - 393
 ACCCACGTCATGACCATCACGAGGGTGTGCTTAACACTCTGTTCCAAACTCTCCGATCCATATCCAATGGAAGTGGAGATTAAATTCAAGCTGTTAATAA 114240

T T K P K E I T P V N P G T S P L L R Y A A N T G G L A A V V L L C L V I F L I 558
 GACTAATACCAAGGAAATTACCCCGTAAACCCGGAACGTACCACTCTACGATATGCCCATGGACCGGAGGGCTTGACAGTAGTACTTTATGTCGTAAATTTAAAT 117480
 C T A K R M R V K A Y R V D K S P Y N Q S M Y Y A G L P V D D F E D S E S T D T 598
 CTGTAACGGCTAACGAATGAGGGTAAAGCTATAGGTAGACAAGTCGGCTATAACCAAAGCTGTTACCGCTGGCTTCAGTGGACGATTCGGAGACTCGGAATCTAAGGATAC 117600
 US ----->---- TRS
 E E E F G H A I G G S H G G S S Y T V Y I D K T R - 623
 GGAAGAAGAGTTGGTAACGGATTGGAGGAGTCACGGGGTTCGAGTTACCGGTGATAGATAAGCCGGTATCACCGAACCAGGGCAACGCCGAGCGTGTAAATTAAATAA 117720
 AAAACAGTACGCTTTATCCGGTGTATGTTAAATTTATTTTTCTATATAAAAGGGATGGGGTGTCAAGGATCTCGTAGGTTCTGGACTCCAAGGGACCCGAGCCCAGGT 117840
 - S R E Y T R P V G L S G C G L Y 165
 69 CGCGTCAAAAAGCCTGTGACAAATTCCCCGGGGTGTATATAATTCCGGCGGGATGCATTATTATCGGGCAGACAATCCAATAGTCGGAGTCGGCGTGTCCAAACAGCACAATA 117960
 A D F L R H C I G G P R O D Y L E A R I C K K N D P S L G I T P L G A H G L C C W Y 125
 TGCCATCAATATCAGCGCTCGCTGAGCTCCAGTCGCTGGATCCGGCATGCTGCTCGGTATAACGATAACCCGCTCTCGGATCGGGAAACCGAGGGTGGACGTGTCT 118080
 A M L I L R E S L E L H S Q P D P M C Q E R L L S L V A R R I P P V S P P R Q R 85
 ATCCCGTGGAGCTCCCCCACATCCGCTCGAAGGGCTCTTGGCTCCCCACAAACTCGGCACGGCTCTCGGGAGGGCTGTGTATAAAAGCCTCGTAGGCTCGGTAGATAATGGC 118200
 D A A L E G C M R G F A D K A R G W L S A C P R E S P S H I F A E Y A E T S I A 45
 GCATAGCGGGGGCGGAAGTACCGTTGGCGCTCGTTCATGAGAGCTGGTGCCTGGGGTATCGTGTGCGAGTAAAGTCGGCATTCACCGCCGGGTGCTCACCGGGATCC 118320
 C L A A A F Y R E A D H M H L A Q H A P T D H R T C Y L G A Y E G G P H E G R S G 5
 GCAGAGATTCAACACAGCCCGTGTGGGGACCCGTGGGTTAATGGAGACGTTCAACCCAGCACCTCTACCCCTATAATCCACAAACCCCACCTCCCCACAGACAGACATCAA 118440
 C L N M 1
 70 AAAAGACACGAGCCAAACCATGTTATTTATAAAACTACCGCCATGGGGGGGGGTATATCATGCCGGCGGGGTTCTGTCGACGTCGTTCAACATCACCCCCCTGCC 118560
 - V G H P P P I D H R R P A E H R R R K W D V G G E G 253
 CAACGGGGTGGCGCTTCGGGCTGTATATCCGGGTTCTGACCCGGTATCGCTCGATGGCTGGCTGGGGCTCTTGGGACGGGGTGTGACCCATECCCGTCTCGATAACA 118680
 L A T L A K A A T Y E A T E A G A Y R E I A R Q P R K P S P T N C G D G D R Y C 213
 ATCGCCGGTACCCGCTGGYTCACAAGAATCGGTCTCCTCTGATTCTCTCCACTATAAGCTCTCCCGTCCGATTCGGCGTGGGAATCCGAAATTCACATCCGA 118800
 D G P V G D P E C S D T S E E S E E V I D F E G D S S E A D S D R F E I V D S 173
 TGGCGTAGACCCATCATCGTCGCTATCGCTTCACCCATCATCAGATACGTCACATTCCAGTGGCTCCATGCAAGGAACATTGGCGCTCAATGACCCGTTCTCCCGAGG 118920
 P T S G D D D S D D E G G D D S V D C E L A S R H L P V N P A E I F G D E G R P 133
 GGGTCGATGGGACACTCCATTAAGTAGCGCTCYGGGTGGGGTGAAGCTGTAAACATTCCTCCAGCAGCTAACGTCGGGACGCTGTGGTGCACCCATTAGATA 119040
 P R H S V E M L T A R L C Q T Q T P O L Q L L E E W C A L T P R Q R T R G W L Y 93
 AAAGTCGAGGCATATGGCGCAATCTCGTGAATCTCGCTGACCGCTCCGACGACTGTGCAATACATAAAAAAACCGTTGATGTCCTCAAGGAGCTGGGGGGATCTCCA 119160
 F D L C I R R L R R S D R T V R G S S Q A T C L F Y R N I D E V L R A P H I E W 53
 ACCCGGAGTACAAACGGCCGGGGCCCCCGACATTCGGGGCTCTGGTCAAGATCCATATCCATCTTCGTTAACATCAACCGATGCCGGGGTTGACTCGGA 119280
 G P T C F A G P G R S T D R G N L P G P A S G Y E M K G N V D V S A G P K S E S 13
 CGAGTCGCCCCCGTAGCGGTGAGGTGACAAATGTCCTGGGGCTGAGTAACTTTCCCTAAACCGACTCGACGCTGTGCGTTATGAAATCGGACGACCCCTGCACAAACAAA 119400
 1
 S D G R T A P S T C F N
 CACACACCAAACGTTACATCTATGAAATAAGCTACTGGGTTAAATGCCATGGGATTCCGGGGGGGAGACCTTCGATTGGGTTGCTTTATAACACCAAAAAAGGGGGGGGCC 119520
 CCGTGTTTTTTTATCACGTCAAATCGATTAAAGCTGGCTCCATTGGAAATATATATCTGTGAAAGCCCGCCACACCCATAAAACCGACATCGGGGAAAC 119640
 origin of DNA replication-----
 GCGCGAACAAAGAACATCTCTCTCTCTATATATATATATATATAGAAAAGAAAGTCGAACCGGTGGTGGACATGCCAAACATGAACACCATACAGTGA 119760
 ----->
 AAAACGGGAAGTGCAGATGCAGATCAAAGAGTGTATCCGATTGGCTACACCAAGACATCGGGACGCCAATTAAACCCCCCCTTTTACCCCCCACCCCATTCACCC 119880
 <-----
 < 'X' >
 CAGGAAGTGCAGACGGGTTACGCTCAGATATGAGTTCTGACTTGTGTTGATAAAATTTTGTGATTCTACACGGTTAGAGAAATTATGGTATAACATCGGGGAAAC 120000
 reiteration R4 -----
 (A' X A' X A' X A' X A' X)
 TACCGCGCCCCCTCCCATCGCGGGTACCGCGCCCTCCCATCGCGGGTACCGCGCCCTCCCATCGCGGGTACCGCGCCCTCCCATCGCGGGTACCGCGCC 120120
 ----->
 TCCCCATCGGGGGGTTACGTGAAACCCACAACCCCGTGTGTTATGGGTTATCGGGGTTCTGTCGCGCTGACATAATCGTGGAGGGTGGTGTATACGCTTGTGAT 120240
 TCGCGAACGTAATGACGACGGAGGGACCCAAACACCGCTGACGTCATTGATTAACATAGATGCCGATGGGTGGAAACACCCGTGTTATAAGATGTTGAGACA 120350

P D V S A L N A R G V L L L S T R D L A F A G A V E Y L G S R L A S A R R R L L 946
 GCCGGAGCTGTCGCTTGAACGCCGGCGTTTATTACTGTGCAACCCGAGACTGGCTTGCCTGGGGCGCTGAGTATCTAGGCTGCCGGTGGCTCTGCCCGGCCGGTTGCT 123600
 V L D A V A L E R W P R D G P A L S Q Y H V Y V R A P A R P D A Q A V V R W P D 986
 GGTGTTGGACGCCGGCTCGAGGGTGCCAGGGATGGACCCGCTTGTCTAGTACGCTGACGCCGGCCGGCGACCCGAGCCGGCTGCTCCGATGCCAGA 123720
 S A V T E G L A R A V F A S S R T F G P A S F A R I E T A F A N L Y P G E Q P L 1026
 CTCGGCGGTCACAGAAGGACTGCCGGCCGCTGTTGATCGTCGCGCACCTTGGCCAGCGAGTTGCTCGTACGAGACTGEGTTGCCAACCTGTACCCGGCGAACACCCCT 123840
 C L C R G G N V A Y T V C T R A G P K T R Y V P L S P R E Y R Q Y V L P G F D G C 1066
 GTGTTGTCGCCGGTGGGAACGTCGACATACCCGGTGTACCCCGCGGCCCAAGACCCGGCTGCCCTGTGCTGCCGTGAATACCCGGCAGTACCGTGTGCCGGGTTTGACGGTTG 123960
 K D L A R Q S R G L G L G A A D F V D E A A H S H R A A H R W G L G A A L R P V 1106
 CAAGGACCTCGGGGACAGTCGGGGCTTGCGGACGGCTTGTGGACGGCGCACATAGCCACCGCGCACAAACCGATGGGCCTGGGTGCCGCTTGTGACCCGT 124080
 F L P E G R R P G A A G P E A G D V P T W A R V F C R H A L L E P D P A A E P L 1146
 CTTCCTCCGGAGGGAGCCGGAGACCCGGGGCCGCCGGGGGAGGCCGAGCTACCCACCTGGCGAGGGTGTGTTGCCGCCACGCCCTGCTGGAAACCCGACCCCTGCCGAGAACCACT 124200
 V L P P V A G R S V A L Y A S A D E A R N A L P P I P R V M W P P G F G A A E T 1186
 CGTGCCTTCAACCGTGGCCGTCGGTGGCCGCTGTATGCGTGGCCGAGGGCTCGGAATGCCCTCCCCCGATTCCTAGAAATGTGGCCGCCGGTTTGCGGCCGGAGAC 124320
 V L E G S D G T R F V F G H H G G S E R P S E T Q A G R Q R R T A D D R E H A L 1226
 GGTGTTGGAGGGAGCGACGGAACACGGTTCGTTGGACACCACGGGGCTCGGAACGCCGAGGGGAGCCGACAGCGGGCACCGCACAGAGAACACGCTT 124440
 E L D D W E V G C E D A W D S E E G G G D D G D A P G S S F G V S I V S V A P G 1266
 GGAGCTGGACGATTGGAGGGTGGGTGTGAAGACCGCTGGACAGCGAGGGGGGGGACGACGGGGGACGGCACGGGGTCATCTTGGGGTAGCCATCGTGTGGGTGGCCCCGGG 124560
 V L R D R R V G L R P A V K V E L L S S S S S S S E D E D D V W G G R G G R S P P 1306
 TGTCGTCGAGACCCCGGGTGGTTTGGCCCGCCGCTCAAGGTGGAGCTGTTGCTCTCGTCCTCCGAGGAGGGACGATGTGTGGGGAGGGCGCGGGGGAGCCCCCCC 124680
 Q S R G - 1310
 GCAGAGTCGGGGGTGACGGAGTCCCTCCCTTCTGAGCGCCACTGGCGCGGACTGTTGTTAATAAAAGCGGAACGGTTTATGAAAAAGTGTCTGTCTGTGCGG 124800
 GCGGGCCAGGGGGGGCTGGTGGACCCCCCCCCGAAAATAACCCCCCCCCGGTTCTGGGCGCCGGAGACCCGGAGGG 124884

Fig. 1. Sequences of the VZV genome and encoded proteins. The DNA sequence is shown as the rightward 5' to 3' strand only. The leftward strand would have an additional C residue at the 3' end and lack a C residue complementary to the G residue at the 3' end of the rightward strand (Davison, 1984). The IR_L-IR_S junction would also be displaced one residue to the left on the leftward strand. Rightward encoded protein sequences are shown in single letter amino acid code above the corresponding DNA sequence. Leftward encoded proteins are shown below the DNA sequence. ORFs are designated by number at the left of the first line containing the amino acid sequence, regardless of coding orientation. The last nucleotide of each AATAAA-related element predicted to function in transcript polyadenylation is indicated by an asterisk above or below the DNA sequence. The locations of the ends of VZV dPyK mRNA are indicated between genes 35 and 37. Four potential TATA signals and three AATAAA-related elements potentially involved in polyadenylation are underlined. The two ATG codons in the untranslated 5' region of this mRNA are marked with asterisks; they are in different frames from the initiation codon for dPyK and are followed by termination codons. A copy of the VZV DNA sequence will be deposited in the EMBL sequence library.

identical in two independent overlapping clones (*Kpn*I c and *Hind*III a). R1 is not located in a recognized region of size variability between virus isolates (Straus *et al.*, 1983).

It is possible that regions other than those containing R2, R3 and R4 are variable in size, as the analyses of VZV isolates would not have identified regions where size differences are small or very infrequent in occurrence. For example, McGeoch *et al.* (1985) described a tract of G:C base pairs in the U_s component of HSV-1 which varies in single base pair steps. Nonetheless, the results described above imply that the VZV genome is not unique in size. The extent of genome size variability among different virus isolates, and the DNA sequences of the reiterations, indicate that the genome may vary in length from just above 124 000 bp to more than 126 000 bp.

The discovery of tandem G + C-rich reiterations in a herpesvirus genome was first made with HSV-1 (Davison & Wilkie, 1981; Watson *et al.*, 1981b), and the role of these sequences in genome size variation was first established with this virus (Davison & Wilkie, 1981). However, it is clear from the HSV-1 sequence data already available that VZV has far fewer reiterations

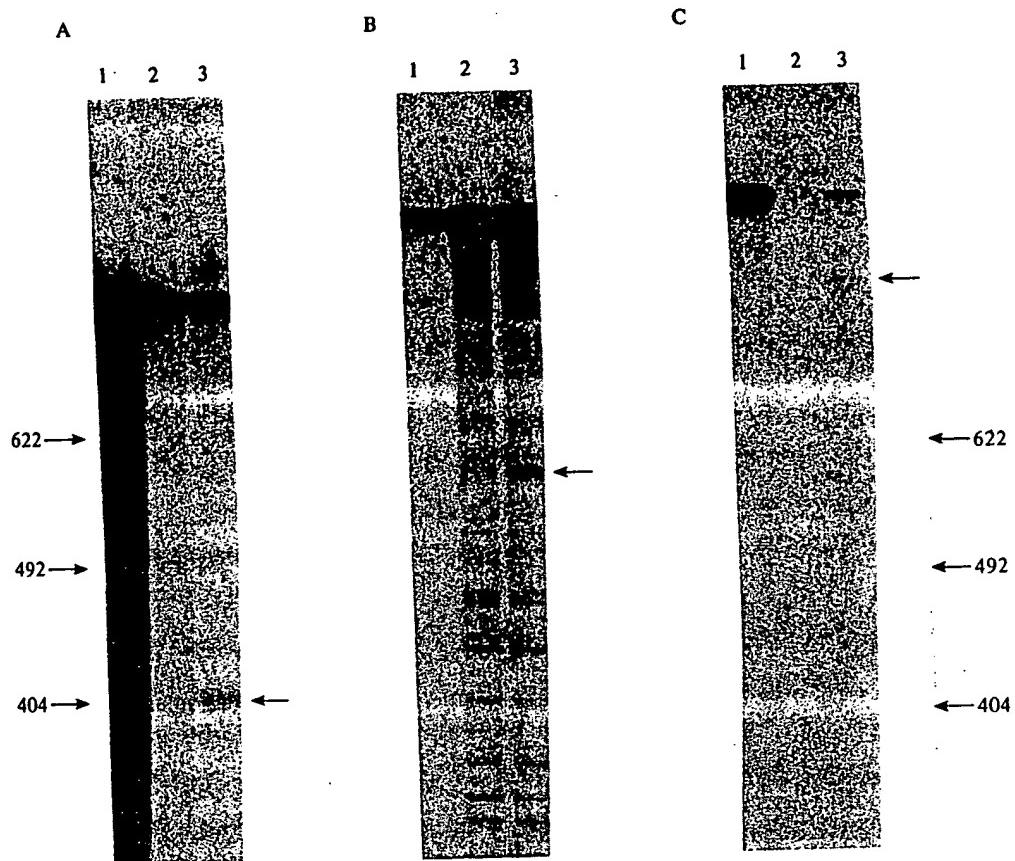


Fig. 4. S1 nuclease analysis of VZV dPyK mRNA. The three end-labelled probes used were (A) 881 nucleotide fragment 5'-labelled at the *Acc*I site and extending leftwards to a *Pst*I site, to detect the 5' end of dPyK mRNA; (B) 1275 nucleotide fragment 3'-labelled at the *Xma*I site and extending rightwards to a *Pst*I site, to detect the 3' end of dPyK mRNA; (C) 1390 nucleotide fragment 5'-labelled at the *Xma*I site and extending leftwards to a *Pst*I site, to detect the 5' end of dPyK mRNA. Each set of three lanes shows (1) untreated probe; (2) S1 nuclease treatment of probe which had been incubated with uninfected cell RNA; (3) S1 nuclease treatment of probe which had been incubated with VZV-infected cell mRNA. Lane 1 contained one-tenth the amount of probe represented in lanes 2 and 3. The sizes of DNA markers in nucleotides are shown on each side of the figure. DNA fragments protected by VZV-infected cell RNA are indicated by arrows.

proposed for some genes whereas no good candidates may be found for others. Thus, the confident identification of a potential promoter for every VZV gene is precluded.

Gene 36 provides a salutary lesson in this respect. The product of this gene has significant homology to HSV-1 deoxypyrimidine kinase (dPyK), and is therefore an excellent candidate for the VZV dPyK detected experimentally (Doberson *et al.*, 1976). The molecular weight of the predicted protein (37815) is in good agreement with the experimentally determined value of 35000 for the monomer subunit (Lopetegui *et al.*, 1983). Fig. 4 shows the results of mapping the ends of the dPyK mRNA, and their locations with respect to the DNA sequence in Fig. 1 are indicated in the region of gene 36. The S1 nuclease results indicate that the 5' end of the mRNA maps 410 bp upstream from the *Acc*I site (Fig. 4, lane A3) and 920 bp upstream from the *Xma*I site (Fig. 4, lane C3), consistent with the function of the sequence TATTAAA underlined in Fig. 1 (64364) as the TATA element. Three similar sequences (TATATTA, TATAAAA and TATAATA) in Fig. 1 (64433, 64469 and 64611) are present between this element and the initiation codon for dPyK. Thus, the location of the 5' end of dPyK mRNA could not have been

predicted from the DNA sequence alone. Similarly, the 3' end of the mRNA might have been expected to map near the AATAAA or ATTAAA elements underlined in Fig. 1 (66125, 65978). In fact, the 3' end is located 580 bp downstream from the *Xma*I site (Fig. 4, lane B3), and is thus defined by the AGTAAA element (65859). This result counsels caution in predicting the 3' ends of other VZV mRNAs; nonetheless, the general degree of confidence was sufficient for proposed elements involved in polyadenylation to be included in Fig. 1 and 2. In any case, there are precedents for the function of AGTAAA in polyadenylation (Donehower *et al.*, 1981; Tamura *et al.*, 1981; Capon *et al.*, 1983). The possibility that the dPyK AGTAAA resulted from a base change in an AATAAA during cloning was ruled out by sequencing this region in an independent clone (*Sst*I h). The element AGTAAA has also been proposed in Fig. 1 and 2 to function in polyadenylation of the transcript from gene 28, which encodes the DNA polymerase.

The dPyK gene is interesting in other respects. Although it encodes a protein with substantial homology to HSV-1 dPyK, as described below, the untranslated 5' region of the mRNA is considerably larger, at 420 bp, than that of the HSV-1 mRNA, at 110 bp (McKnight, 1980). Moreover, the untranslated 5' region of the VZV gene contains two ATG codons in different reading frames from the initiation codon for dPyK. Each of these is followed within a short distance by a termination codon and, in view of the work of Kozak (1984), this structural aspect may affect translational efficiency of the mRNA by requiring reinitiation for expression of dPyK. As the only VZV gene promoter identified experimentally to date is that for the dPyK gene, responsible comparisons between HSV-1 and VZV promoter regions may be made only for this gene. Little similarity was detected between regions upstream of the TATA elements. The differences between the structures of the promoters and untranslated 5' mRNA regions of the VZV and HSV-1 dPyK genes suggest that transcriptional and translational control might differ considerably between the two genes.

The locations of potential polyadenylation sites near the 3' ends of ORFs are summarized in Fig. 2. Many genes apparently possess unique polyadenylation sites, whereas others are present in 3'-coterminal families containing up to four genes. The 3'-coterminal gene arrangement has been well-characterized in HSV-1, where sets of genes are expressed as mRNAs with unique 5' ends but shared 3' ends (for review, see Wagner, 1985). Thus, the mRNAs expressed from genes towards the 5' end of a 3'-coterminal family contain extensive untranslated 3' regions. The VZV genome contains 216 AATAAA elements, but only 48 unique potential polyadenylation sites are predicted in Fig. 1 and 2. Moreover, seven of these sites contain ATTAAA rather than AATAAA, and two contain AGTAAA. It is possible that the presence of some AATAAA-related elements close to the ends of ORFs is merely fortuitous, and thus some genes shown with unique polyadenylation sites in Fig. 1 and 2 actually may be members of 3'-coterminal families. This comment applies particularly to VZV mRNAs predicted to have unique 3' ends whose HSV-1 counterparts are polyadenylated as part of a 3'-coterminal family. For example, HSV-1 transcripts corresponding to VZV ORFs 40 and 41 are 3'-coterminal (Costa *et al.*, 1981), as are those corresponding to ORFs 43 and 44 (Costa *et al.*, 1984). Ostrove *et al.* (1985) recently reported a transcript map for the VZV genome, based on Northern blot analysis using relatively large cloned DNA fragments. The arrangement of VZV mRNAs suggested by this approach correlates well with that deduced from the DNA sequence in some regions, but there are apparent discrepancies in others. Confirmation of the transcript map and resolution of these differences must await the mapping of mRNA termini.

Although identified overlapping polypeptide-coding regions are not extensive, the VZV genome shows considerable economy in its gene layout. Almost the entire sequence encodes virus proteins, and it is likely that many regions involved in control of gene expression are located in the coding regions of adjacent genes. However, there are four notable regions for which no protein products are predicted. One is located at the left end of the L segment and contains TR_L. It may contain sequences which promote cleavage of full-length genomes from concatemers produced during DNA replication (Davison, 1984). The second region, also about 600 bp in size, is located at the right end of the L segment and contains IR_L. Part of its function is likely to be as a promoter for gene 61. The third region is about 1400 bp in size and is located

between genes 60 and 61. It contains an unusual direct repeat of 88 bp, with three mismatches, separated by 24 bp (102020–102219 in Fig. 1; A indicates each repeat). The function of this structure is unknown; it could form part of a control element for gene 60. The fourth region is about 1400 bp in size and is present twice in the genome: in IR_S between genes 62 and 63 and in TR_S between genes 70 and 71. It contains the reiteration R4 and the promoters for the two genes on either side. It also contains a palindrome (Davison & Scott, 1985) which forms part of a functional origin of DNA replication (Stow & Davison, 1986). In view of the compact arrangement of the rest of the genome, these four regions are likely to have important functions. Although no protein products have been assigned to them, it is possible that some contain small coding exons or perhaps larger non-coding exons. Alternatively, they may encode functional RNAs which are not expressed as proteins. A third alternative is that they may encode no RNA or protein species, but are sites for specific recognition during the virus life-cycle. This is certainly the case for the origin of DNA replication, and probably for sequences at the left end of the genome potentially involved in DNA maturation.

VZV gene function

Comparison of the proposed arrangement of VZV genes in Fig. 2 with published HSV-1 transcript mapping data (for review, see Wagner, 1985) indicates that both viruses have a similar gene layout. This view was confirmed by available HSV-1 sequence data, and allowed the functions of several VZV genes to be assigned on the basis of primary amino acid sequence homology of their products to HSV-1 proteins. These conclusions, and the precise locations of VZV genes and molecular weights of their primary translation products, are summarized in Table 1. Genes encoding glycoproteins, homologues of HSV-1 immediate-early proteins, and proteins with extreme properties of hydrophobicity, hydrophilicity, charge or amino acid composition are also indicated. All but three of the functional assignments were made on the basis of HSV-1 gene location and confirmed by amino acid sequence homology with HSV-1 proteins. The dUTPase was assigned on the basis of the location of the HSV-1 gene reported by Preston & Fisher (1984). The thymidylate synthetase and protein kinase genes were located on the basis of amino acid sequence homology of their products to proteins of known function in the NBRF protein database. Approximately 30 VZV proteins are homologous to proteins predicted from the complete EBV sequence determined by Baer *et al.* (1984); the implications of this result in predicting the functions of EBV genes will be discussed elsewhere (A. J. Davison & P. Taylor, unpublished data).

Fig. 5 shows examples of homology between VZV and HSV proteins displayed by optimal alignment of predicted amino acid sequences. Fig. 5(a) shows a comparison of the product of VZV gene 18 with the small subunit of the HSV-2 ribonucleotide reductase; these proteins are highly conserved. The lower degree of homology between the VZV and HSV-1 dPyKs shown in Fig. 5(b) is in accord with the DNA hybridization data of Davison & Wilkie (1983), who were able to detect conservation of the ribonucleotide reductase gene but not of the dPyK gene. The degree of homology shown in Fig. 5(c) between the product of VZV gene 5 and the potential HSV-1 membrane protein reported by Debroy *et al.* (1985) is about the same as that observed for the dPyKs. However, several pairs of genes are less conserved than this, and only specific regions of the proteins were detected as being conserved by this approach. Fig. 5(d) shows the conservation of a region towards the carboxy termini of the glycoprotein product of VZV gene 14 and HSV-1 glycoprotein C (gC). Although the homology in this region is significant, the major part of each protein is divergent. The divergent region of the VZV protein contains a repeated amino acid sequence coded by reiteration R2.

Most glycoprotein genes encode primary translation products with distinct characteristics: a hydrophobic signal sequence near the amino terminus for translation of the mRNA on membrane-bound ribosomes (Blobel, 1980) and a more extensive hydrophobic region followed by basic residues near the carboxy terminus for anchoring the protein in the membrane (Tomita & Marchesi, 1975). The VZV genome contains five such genes: 14, 31, 37, 67 and 68. Genes 14, 31, 37 and 68 are counterparts of identified HSV-1 glycoprotein genes, as shown in Table 1. Gene 67 is the counterpart of an HSV-1 gene whose predicted product also has the

Table 1. Properties of proteins coded by predicted VZV genes

Gene	Start*	Stop†	Residues	Mol. wt.‡	Extreme properties§	Function
1	915	592	108	12103	Hydrophobic (C)	
2	1134	1847	238	25983		
3	2447	1911	179	19149		
4	4141	2786	452	51540	Hydrophilic (N)	
5	5274	4255	340	38575	Hydrophobic	<u>Homologue of HSV-1 IE63¹</u>
6	8577	5329	1083	122541		
7	8607	9383	259	28245		
8	10667	9480	396	44816		dUTPase
9	11009	11914	302	32845	Hydrophilic	<i>trans</i> -inducing factor ²
10	12160	13389	410	46573		
11	13590	16046	819	91825	Hydrophilic & acidic (N)	
12	16214	18196	661	74269		
13	18441	19343	301	34531		Thymidylate synthetase ³
14	21113	19434	560	61350		Glycoprotein (gpV); homologue of HSV-1 gC ⁴
15	22478	21261	406	44522	Hydrophobic	
16	23794	22571	408	46087		
17	24149	25513	455	51365		
18	26493	25576	306	35395	Acidic	Small subunit of ribonucleotide reductase ⁵
19	28845	26521	775	86823		Large subunit of ribonucleotide reductase ⁶
20	30475	29027	483	53969		
21	30759	33872	1038	115774		
22	34083	42371	2763	306325		
23	43138	42434	235	24416	Hydrophilic; S, T, Q-rich	
24	44021	43215	269	30451	Hydrophobic (C)	
25	44618	44151	156	17460	Hydrophilic; acidic (N)	
26	44506	46260	585	65692		
27	46127	47125	333	38234	Hydrophilic & basic (N)	
28	50636	47055	1194	134041		DNA polymerase ⁷
29	50857	54468	1204	132133		Major DNA-binding protein ⁷
30	54651	56960	770	86968		
31	57008	59611	868	98062		Glycoprotein (gpII); homologue of HSV-1 gB ⁸
32	59766	60194	143	15980	Hydrophilic & acidic	
33	62138	60324	605	66043		
34	63910	62174	579	65182		
35	64753	63980	258	28973	Basic	
36	64807	65829	341	37815		Deoxypyrimidine kinase ⁹
37	66074	68596	841	93646		Glycoprotein (gpIII?); homologue of HSV-1 gH ¹⁰
38	70293	68671	541	60395		
39	70633	71352	240	27078	Hydrophobic	
40	71540	75727	1396	154971		Major capsid protein ¹¹
41	75847	76794	316	34387		
42	78038	76854	395	82752		
43	82593	81538	352	(spliced)		
44	78170	80197	676	73905		
45	80360	81448	363	40243		
46	82719	83315	199	22544	Hydrophilic & acidic	
47	83168	84697	510	54347		
48	84667	86319	551	61268		Exonuclease ¹²
49	86226	86468	81	8907	Hydrophilic	
50	87882	86578	435	48669	Hydrophobic	
51	87881	90385	835	94370		
52	90493	92805	771	86343		
53	93850	92858	331	37417		
54	95984	93678	769	86776		
55	95996	98638	881	98844		
56	98568	99299	244	27166	S, T-rich	
57	99626	99414	71	8079	Hydrophilic & basic	
58	100272	99610	221	25093	Hydrophilic & basic	
59	101219	100305	305	34375		
60	101649	101173	159	17616	Acidic	
61	104485	103085	467	50913	Hydrophilic	
62	109133	105204	1310	139989		Homologue of HSV-1 IE175 ¹³
63	110581	111414	278	30494	Hydrophilic & acidic	Homologue of HSV-1 IE68 ¹⁴
64	111565	112104	180	19868		
65	112640	112335	102	11436	Hydrophobic (C)	
66	113037	114215	393	43677		Protein kinase ¹⁵

characteristics of a glycoprotein (McGeoch *et al.*, 1985). The glycoprotein products of VZV genes 31, 67 and 68 have been identified unequivocally: they encode gpII (Keller *et al.*, 1986), gpIV (Davison *et al.*, 1985) and gpI (Ellis *et al.*, 1985), respectively, according to the recently established VZV glycoprotein nomenclature described by Davison *et al.* (1986). Thus, there are two probable VZV glycoprotein genes (14 and 37) whose products have not yet been identified, and one antigenically defined major glycoprotein (gpIII) whose gene has not yet been mapped. In the absence of additional data, the product of gene 37 has been tentatively assigned as gpIII? in Table 1, and that of gene 14 has been proposed as gpV, a previously undetected minor glycoprotein.

The VZV genome also contains four genes (5, 15, 39, 50) which encode particularly hydrophobic proteins. Mutations in the HSV-1 counterpart of gene 5 result in a syncytial plaque morphology, and it has been suggested that the product of this gene is a membrane protein (Debroy *et al.*, 1985). Thus, it is possible that VZV gene 5, and perhaps genes 15, 39 and 50, encode membrane-associated proteins.

Although the HSV-1 and VZV gene arrangements are similar, there are limited regions of significant difference. The most extensive encompasses the S segment, which in HSV-1 contains 13 unique genes (McGeoch *et al.*, 1985, 1986a) and in VZV contains only seven. The relationship between the S segments of VZV and HSV-1 has been discussed separately (Davison & McGeoch, 1986). In summary, each VZV gene has a homologue in HSV-1, but the remaining six HSV-1 genes have no counterparts in VZV. The 'missing' genes include the immediate-early gene which encodes HSV-1 IE12, and the gene which encodes glycoprotein D. These regions of the two genomes differ substantially in gene layout, but they are clearly related, and a scheme has been proposed for their descent from the S segment of an ancestral herpesvirus by expansion and contraction of the inverted repeats. Two other regions of difference between the VZV and HSV-1 genomes are present at the ends of the L segment. These regions have been sequenced in HSV-1 (L. J. Perry & D. J. McGeoch, personal communication), and appear not to contain homologues to VZV gene 1 and perhaps gene 2, at the left end, and gene 61 at the right end. Also, the inverted repeats flanking U_L are much larger in HSV-1, at approximately 9000 bp, than they are in VZV (88.5 bp), and the single gene thus far identified from the HSV-1 sequence of this region specifies a spliced immediate-early mRNA encoding IE110 (L. J. Perry, F. J. Rixon & D. J. McGeoch, personal communication). At the present stage of analysis, no homologue of this gene has yet been detected in VZV. Thus, the differences in gene arrangement between VZV and HSV-1 in the S segment and at the ends of the L segment result in VZV apparently lacking homologues to two of the five HSV-1 immediate-early genes: those encoding IE12 and IE110. There is only one other region for which sufficient HSV-1 data are available to indicate a difference in gene layout between VZV and HSV-1. It is almost certain, from several lines of evidence, that HSV-1 lacks a homologue to VZV gene 13, which encodes a protein with a

67	114496	115558	354	39362	Glycoprotein (gpIV); homologue of HSV-1 US7 ¹⁴
68	105808	117676	623	69953	Glycoprotein (gpI); homologue of HSV-1 gE ¹⁴
69	118332	117793	180	19868	Homologue of HSV-1 IE68 ¹⁴
70	119316	118483	278	30494	Hydrophilic & acidic
71	120764	124693	1310	139989	Homologue of HSV-1 IE175 ¹³

* Location in rightward strand of first base of initiating ATG codon. All except 14, 31 and 68 refer to the first ATG in the ORF (see text).

† Location in rightward strand of last base of the codon preceding termination codon.

‡ All predicted mol. wt. values except those of the primary translation products of genes 14, 31 and 68 (see text), were calculated from the amino acid sequence commencing at the first in-frame ATG in the ORF.

§ (N) and (C) indicate that the extreme property is limited to a region towards the amino or carboxy terminus, respectively.

¶ References are given in the superscripts to HSV sequence data, or other sequence data for genes 13 and 66, which confirm the assignment of VZV gene function. ¹L. J. Perry & D. J. McGeoch, personal communication. ²Dalrymple *et al.* (1985). ³A. J. Davison & R. W. Honess, unpublished data. ⁴Draper *et al.* (1984). ⁵Y. Nikas & J. B. Clements, personal communication.

⁶McLauchlan & Clements (1983). ⁷Quinn & McGeoch (1985). ⁸Bzik *et al.* (1984). ⁹McKnight (1980). ¹⁰U. Gompels & A. C. Minson, personal communication; McGeoch & Davison (1986b). ¹¹A. J. Davison & J. E. Scott, unpublished data. ¹²Draper *et al.* (1984); McGeoch *et al.* (1986b). ¹³McGeoch *et al.* (1986a). ¹⁴McGeoch *et al.* (1985). ¹⁵McGeoch & Davison (1986a).

(a)

MDQKDCSHPFYRPECPIINNLKALISIHWLLESQVITIHDYQYDLCIETEDELIFYRFITPTPLSAADDLVNVNLGSLTOLFSOKDHHYI[KOECLEVVHAR
 102 VVSVSIOLMLFGRGDESLHVQYVWVTEINNPSIQQKVONLVEKVRUNPVAEYKILKILIGCIPYVSSAAIAYLKNNGLFVVTQDQYKULISRDEAIHTSASCCLYNNYVPEKPAITRIHOLPSEAY
 127 VVNIQOLVLFHNNDOARRAYVARTINHPAIRVKVDMLEARVRECDSPVNPENFTLRLTEGVPPYASPAIAYLRTNLLRVTCOSWDLISRDEAVHTTASCYIHNHYLGDHAKPEARVYRLPSEAY
 226 KIECAFLKSHAPKTHLWVNUAATQYVVKPSADRLSAINVPKLNFNTPPUSDPFLAFMADKNTNYYEIIISTSYAGTVINDE
 253 DLEIGFIRSOQPTDSILSPGALAATENYVPSADRLLCLIHMOPLSYAPAPUASPPSLSMSTUKHTMFECRKSTSAYGAVVNDL

(b)

I6-TURTOVKKGVHLKYLIGAYGIGKTTAEEFPHIATTPNPIILLGEPLSYWRNLAGEDAICGYGOTRILNGUVSPEDAORLTAF
 1 HASVPCPHQHSAFUAQSKHGNHNTAJPQRHNUQKATEVRLEQKPTILHRYTDQPHMGKTTTTDILVALGSKUUVVYVPEPNTWVRLGABETIANVTTQHRLDUGESIAGDAAVVNTSA
 90 USIYPCPHAIHAKISALHDTSTSULQVNKEPYKIMLSKIIPIASTICPPSLKYLQVLMPSAIIJGQFLIFTLPAPPPGTNLVVCVTSVSLPSHLSRSVSKRABGETVNLPPVNLKVVYIMLINTI
 125 QITMGKHPYAVTDAVLAPHIGGEAGSSHAPPALTLIURPIIAALLCPAARYLMSHTPOAVLAPVALIPTPLPGTNIVLGAPEDRHIDRLAKRQHGERLDLAMLAIRVYVGLLANTV
 215 PLKTNHHAGWNTSFCHUVPKLQLKSECILKREVPGTEUTLPFLVPLKLIPLCGEGFNLIPPLWAMHETLSNCNSHSMPVLSLEOTQHAAQELKTLLPOMTPANMSSGANILKELY
 248 YLOGGGSWRELDNGULSAAVPMUGALEPUSNAGPKPHIGUTLPLETKEPELLAPNGULYNYVAVALUVLAKRLRPHVYILOYDOSPAGCRDALLQTSCHMVOTHVTPGSIPTICDLARTF

334 MAVQINTS

369 AREMGLEAN

(c)

I MAQGIKTEHPIIMCLSGHAAVTLWDT ARVKPEIECVYA TTVINCGPVWVGYNNSLIVYTFVNHSTFLDGLSGYDYSCRENLSGOTMVKAISTPLHDKIRIVLGTRNCHAYFWCVOA
 1 MIA VRSLQHNSSTVVLITAYGLVNWYTVFGASPLHICIAVHPTGNNHNTALWVWNKHNTLFLPGATHPPNGWNNHAIISYANLIAGRVVPFQVPPDATHKRNIVHEAVNCLETLYTRV
 123 KIIYAWVYVGYLQHRIIRHOPPFRSCELISPTSISLNVTRVVISNLLQYHYYKLAHLICUDVSHRRKDKMSKVENADPISFLY NHKGVTLLHLLLEVIAHISGCIULLTLCVAYTPCALLY
 125 LVVVGWFLYLAVALHQRKCHGPGVVSFAHKMVAPATYLANYAGRIVSVLQYPTKITHLCLCELSYORQNLVOLFETDPUFLYHRPAIGVIVGCELIVRFAVGLIVGTAISRGACAITY
 247 TYVIRILAWVVCTLAIVELISVVBWPKPTKDNNLHNHNG GIRCICTTTACATVMSGLAKCFYIVIAIAVVIPIHYYEYORVQVSLFGESENOKH
 249 LFLTITTCYVSTIGLTELTYCLLRRGPAPKNAKAAAGRSKGGLSGVCGKCCSILSGCIAIRLCYIAVVGVVLVALHYEQIIORRFLDV

(d)

301 RYVNTTKNVISEHSIT VTTYYRPNITVUVGDPPVLTGOTYAAVCNVSKVYYPHSVVRVWTSHPGNIKKNFTDIAOEYANGLFSYVSAVRIPOOKGHWYPPPQIQCWVWIRDGVSHMKYSAV
 321 HYVYLANGRMDSPHEYGTWVVRVWFRPPSLTLQPHAVMEQGPFKATCTAAAYTPRNPVFDWFEDDRQVNPQGIDTOTHEHDPGFTTVST TSEAVGGVQPPRTFTCOMTWHRSUTFSKRMATG
 424 TPWVYPPPNVSIGIIDGHIVCTACVPGVUVHFWVWNHSPINHENSEITGVCQDONKRFVNMNOSSCPTSELUGPITYSCHLGDYPKKP PFSAVVTTDASTYATTFSVVAIIGVISILGTL
 366 LALVLPRPTITMEFGVRIIVVCTAGCVPLG VTFAMFLGDDPSPAAKSAVTAQESCONIPLATVRSVTPISYD YSEYICRLTGYPAGIPVLEHNGSHOPPPRUPTEROVIETEWVGIGIGVLA
 547 LIAVIAATLICRCOS
 490 VLVVTATIVVVRTSQSRHRR

Fig. 5. Optimal alignment displays of the predicted (single-letter) amino acid sequences of (a) the product of VZV gene 18 and the small subunit of HSV-2 ribonucleotide reductase (McLauchlan & Clements, 1983); (b) the product of VZV gene 36 and HSV-1 strain 17 dPyK (D. J. McGeoch, personal communication); (c) the product of VZV gene 5 and the potential HSV-1 membrane protein described by Debroy *et al.* (1985); (d) the product of VZV gene 14 and HSV-1 gC (Draper *et al.*, 1984). In each example the VZV sequence is shown above the HSV sequence.

remarkable degree of homology to prokaryotic and eukaryotic thymidylate synthetases (A. J. Davison & R. W. Honess, unpublished data). Instead, Frink *et al.* (1983) have shown that this region of the HSV-1 genome contains a small gene which is present in a 3'-coterminal family with the gC gene (the counterpart of VZV gene 14) and is thus in the opposite relative orientation from VZV gene 13. Thus, although VZV and HSV-1 are very similar in gene layout in the L segment, this discovery enhances the possibility that one or more other local differences may exist.

The VZV and HSV-1 genomes also differ in another functional aspect. The region between the HSV-1 DNA polymerase and major DNA-binding protein genes contains a large palindrome (Gray & Kaerner, 1984; Weller *et al.*, 1985; Quinn & McGeoch, 1985) which forms part of an origin of DNA replication (Weller *et al.*, 1985). This origin is termed *ori_L* to distinguish it from *ori_S* in TR_S and IR_S. Plasmids containing the corresponding region of the VZV genome do not contain a palindrome and do not function as origins (Stow & Davison, 1986). Comparison of cloned and virion DNA fragments (data not shown) has ruled out the

possibility that a similar palindrome might have been deleted during cloning, as occurs in clones containing HSV-1 *ori_L*. Therefore, although VZV has an origin corresponding to HSV-1 *ori_S* (110087 to 110350 and 119547 to 119810 in Fig. 1; Stow & Davison, 1986), it does not possess a second functional origin in a location equivalent to that of HSV-1 *ori_L*. However, it cannot be ruled out that VZV has a second origin elsewhere in the genome which may have been deleted on cloning.

Significance of the VZV sequence

The DNA sequence provides a firm foundation on which to build a detailed understanding of VZV infection at the molecular level. This knowledge may be applied in the development of effective treatments for the diseases caused by this virus. The sequence has also given the first complete view of gene layout in a member of the *Alphaherpesvirinae*, and has allowed our knowledge of the proposed functions of VZV genes to increase from almost nothing to equal that of HSV-1. Clearly, the sequence will be important in determining the functions of the majority of VZV genes whose role in virus growth is at present unknown. The way in which data from one herpesvirus may be so usefully applied to another thus encourages herpesvirologists to cultivate a more catholic approach towards the family of viruses they study.

The authors are grateful to John Subak-Sharpe for continuing interest in this work, to Duncan McGeoch for advice and discussion, to Philip Taylor for extensive help with computer programming and to Catherine Joyce for supplying the *E. coli* strain which overproduces the Klenow fragment of DNA polymerase I. Thanks are due also to Duncan McGeoch, Frazer Rixon, Lise Perry, Nigel Stow, Mike Dalrymple, Yannis Nikas, Ursula Gompels and Ed Wagner for making available unpublished HSV-1 data, and to John Subak-Sharpe and Duncan McGeoch for critically reading the manuscript.

REFERENCES

- BAER, R., BANKIER, A. T., BIGGIN, M. D., DEININGER, P. L., FARRELL, P. J., GIBSON, T. J., HATFULL, G., HUDSON, G. S., SATCHEWELL, S. C., SEGUIN, C., TUFFNELL, P. S. & BARRELL, B. G. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature, London* **310**, 207-211.
- BERK, A. J. & SHARP, P. A. (1978). Spliced early mRNAs of simian virus 40. *Proceedings of the National Academy of Sciences, U.S.A.* **75**, 1274-1278.
- BIGGIN, M. D., GISSON, T. J. & HONG, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 3963-3965.
- BLOBEL, G. (1980). Intracellular protein topogenesis. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 1496-1500.
- BLUMENTHAL, R. M., RICE, P. J. & ROBERTS, R. J. (1982). Computer programs for nucleic acid sequence manipulation. *Nucleic Acids Research* **10**, 91-101.
- BZIK, D. J., FOX, B. A., DELUCA, N. A. & PERSON, S. (1984). Nucleotide sequence specifying the glycoprotein gene, gB, of herpes simplex virus type 1. *Virology* **133**, 301-314.
- CAPON, D. J., CHEN, E. Y., LEVINSON, A. D., SEEBURG, P. H. & GOEDDEL, D. V. (1983). Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. *Nature, London* **302**, 33-37.
- CASEY, T. A., RUYECHAN, W. T., FLORA, M. N., REINHOLD, W., STRAUS, S. E. & HAY, J. (1985). Fine mapping and sequencing of a variable segment in the inverted repeat of varicella-zoster virus DNA. *Journal of Virology* **54**, 639-642.
- CHOU, J. & ROIZMAN, B. (1985). Isomerization of herpes simplex virus 1 genome: identification of the *cis*-acting and recombination sites within the domain of the *a* sequence. *Cell* **41**, 803-811.
- CORDEN, J., WASYLYK, B., BUCHWALDER, A., SASSONE-CORSI, P., KEDINGER, C. & CHAMBON, P. (1980). Promoter sequences of eukaryotic protein-coding genes. *Science* **209**, 1406-1414.
- COSTA, R. H., DEVI, B. G., ANDERSON, K. P., GAYLORD, B. H. & WAGNER, E. K. (1981). Characterization of a major late herpes simplex virus type 1 mRNA. *Journal of Virology* **38**, 483-496.
- COSTA, R. H., DRAPER, K. G., KELLY, T. J. & WAGNER, E. K. (1984). An unusual spliced herpes simplex virus type 1 transcript with sequence homology to Epstein-Barr virus DNA. *Journal of Virology* **54**, 317-328.
- DALRYMPLE, M. A., McGEOCH, D. J., DAVISON, A. J. & PRESTON, C. M. (1985). DNA sequence of the herpes simplex virus type 1 gene whose product is responsible for transcriptional activation of immediate early promoters. *Nucleic Acids Research* **13**, 7865-7879.
- DAVISON, A. J. (1983). DNA sequence of the U_S component of the varicella-zoster virus genome. *EMBO Journal* **2**, 2203-2209.
- DAVISON, A. J. (1984). Structure of the genome termini of varicella-zoster virus. *Journal of General Virology* **65**, 1969-1977.

- DAVISON, A. J. & McGEOCH, D. J. (1986). Evolutionary comparisons of the S segments in the genomes of herpes simplex virus type 1 and varicella-zoster virus. *Journal of General Virology* **67**, 597-611.
- DAVISON, A. J. & SCOTT, J. E. (1983). Molecular cloning of the varicella-zoster virus genome and derivation of six restriction endonuclease maps. *Journal of General Virology* **64**, 1811-1814.
- DAVISON, A. J. & SCOTT, J. E. (1985). DNA sequence of the major inverted repeat in the varicella-zoster virus genome. *Journal of General Virology* **66**, 207-220.
- DAVISON, A. J. & WILKIE, N. M. (1981). Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2. *Journal of General Virology* **55**, 315-331.
- DAVISON, A. J. & WILKIE, N. M. (1983). Location and orientation of homologous sequences in the genomes of five herpesviruses. *Journal of General Virology* **64**, 1927-1942.
- DAVISON, A. J., WATERS, D. J. & EDSON, C. M. (1985). Identification of the products of a varicella-zoster virus glycoprotein gene. *Journal of General Virology* **66**, 2237-2242.
- DAVISON, A. J., EDSON, C. M., ELLIS, R. W., FORGHANI, B., GILDEN, D., GROSE, C., KELLER, P. M., VAFAI, A., WRÓBLEWSKA, Z. & YAMANISHI, K. (1986). New common nomenclature for glycoprotein genes of varicella-zoster virus and their glycosylated products. *Journal of Virology* **57**, 1195-1197.
- DEBROY, C., PEDERSON, N. & PERSON, S. (1985). Nucleotide sequence of a herpes simplex virus type 1 gene that causes cell fusion. *Virology* **145**, 36-48.
- DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387-395.
- DOBERSON, M. J., JERKOFSKY, M. & GREER, S. (1976). Enzymatic basis for the selective inhibition of varicella-zoster virus by 5-halogenated analogs of deoxycytidine. *Journal of Virology* **20**, 478-486.
- DONEHOWER, L. A., HUANG, L. & HAGER, G. L. (1981). Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. *Journal of Virology* **37**, 226-238.
- DRAPER, K. G., COSTA, R. H., LEE, G. T.-Y., SPEAR, P. G. & WAGNER, E. K. (1984). Molecular basis of the glycoprotein-C-negative phenotype of herpes simplex virus type 1 macroplaque strain. *Journal of Virology* **51**, 578-585.
- DUMAS, A. M., GEELEN, J. L. M. C., MARIS, W. & VAN DER NOORDAA, J. (1980). Infectivity and molecular weight of varicella-zoster virus DNA. *Journal of General Virology* **47**, 233-235.
- DUMAS, A. M., GEELEN, J. L. M. C., WESTSTRATE, M. W., WERTHEIM, P. & VAN DER NOORDAA, J. (1981). *Xba*I, *Pst*I and *Bgl*II restriction endonuclease maps of the two orientations of the varicella-zoster virus genome. *Journal of Virology* **39**, 390-400.
- ECKER, J. R. & HYMAN, R. W. (1982). Varicella-zoster virus DNA exists as two isomers. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 156-160.
- ELLIS, R. W., KELLER, P. M., LOWE, R. S. & ZIVIN, R. A. (1985). Use of a bacterial expression vector to map the varicella-zoster virus major glycoprotein gene, gC. *Journal of Virology* **53**, 81-88.
- FITZGERALD, M. & SHENK, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* **24**, 251-260.
- FRINK, R. J., EISENBERG, R., COHEN, G. & WAGNER, E. K. (1983). Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. *Journal of Virology* **45**, 634-647.
- GAROFF, H. & ANSORGE, W. (1981). Improvements of DNA sequencing gels. *Analytical Biochemistry* **115**, 450-457.
- GILDEN, D. H., SHTRAM, Y., FRIEDMANN, A., WELLISH, M., DEVLIN, M., FRASER, N. & BECKER, Y. (1982). The internal organization of the varicella-zoster virus genome. *Journal of General Virology* **60**, 371-374.
- GRAY, C. P. & KAERNER, H. C. (1984). Sequence of the putative origin of replication in the UL region of herpes simplex virus type 1 ANG DNA. *Journal of General Virology* **65**, 2109-2119.
- HANAHAN, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**, 557-580.
- JOYCE, C. M. & GRINDLEY, N. D. F. (1983). Construction of a plasmid that overproduces the large proteolytic fragment (Klenow fragment) of DNA polymerase I of *Escherichia coli*. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 1830-1834.
- KELLER, P. M., DAVISON, A. J., LOWE, R. S., BENNETT, C. D. & ELLIS, R. W. (1986). Identification and structure of the gene encoding gpII, a major glycoprotein of varicella-zoster virus. *Virology* **152**, 181-191.
- KINCHINGTON, P. R., REINHOLD, W. C., CASEY, T. A., STRAUS, S. E., HAY, J. & RUYECHAN, W. T. (1985). Inversion and circularization of the varicella-zoster virus genome. *Journal of Virology* **56**, 194-200.
- KOZAK, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Research* **12**, 857-872.
- KUMAR, A. & LINDBERG, U. (1972). Characterization of messenger ribonucleoprotein and messenger RNA from KB cells. *Proceedings of the National Academy of Sciences, U.S.A.* **69**, 681-685.
- KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105-132.
- LOPETEGUI, P., MATSUNAGA, Y., OKUNO, T., OGINO, T. & YAMANISHI, K. (1983). Expression of varicella-zoster virus-related antigens in biochemically transformed cells. *Journal of General Virology* **64**, 1181-1186.
- LUDWIG, H., HAINES, H. G., BISWAL, N. & BENYESH-MELNICK, M. (1972). The characterization of varicella-zoster virus DNA. *Journal of General Virology* **14**, 111-114.
- McGEOCH, D. J. & DAVISON, A. J. (1986a). Alphaherpesviruses possess a gene homologous to the protein kinase gene family of eukaryotes and retroviruses. *Nucleic Acids Research* **14**, 1765-1777.
- McGEOCH, D. J. & DAVISON, A. J. (1986b). DNA sequence of the herpes simplex virus type 1 gene encoding glycoprotein gH, and identification of homologues in the genome of varicella-zoster virus and Epstein-Barr virus. *Nucleic Acids Research* **14**, 4281-4292.

- McGEOCH, D. J., DOLAN, A., DONALD, S. & RIXON, F. J. (1985). Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type 1. *Journal of Molecular Biology* **181**, 1-13.
- McGEOCH, D. J., DOLAN, A., DONALD, S. & BRAUER, D. H. K. (1986a). Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Research* **14**, 1727-1745.
- McGEOCH, D. J., DOLAN, A. & FRAME, M. C. (1986b). DNA sequence of the region in the genome of herpes simplex virus type 1 containing the exonuclease gene and neighbouring genes. *Nucleic Acids Research* **14**, 3435-3448.
- MCKNIGHT, S. L. (1980). The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. *Nucleic Acids Research* **8**, 5949-5964.
- MC LAUCHLAN, J. & CLEMENTS, J. B. (1983). DNA sequence homology between two co-linear loci on the HSV genome which have different transforming abilities. *EMBO Journal* **2**, 1953-1961.
- MARTIN, J. H., DOHNER, D. E., WELLINGHOFF, W. J. & GELB, L. D. (1982). Restriction endonuclease analysis of varicella-zoster vaccine virus and wild-type DNAs. *Journal of Medical Virology* **9**, 69-76.
- MESSING, J. (1979). A multi-purpose cloning system based on the single-stranded DNA bacteriophage M13. *Recombinant DNA Technical Bulletin* **2**, 43-48.
- MESSING, J. & VIEIRA, J. (1982). A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. *Gene* **19**, 269-276.
- MISHRA, L., DOHNER, D. E., WELLINGHOFF, W. J. & GELB, L. D. (1984). Physical maps of varicella-zoster virus DNA derived with 11 restriction enzymes. *Journal of Virology* **50**, 615-618.
- MOUNT, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Research* **10**, 459-472.
- MURCHIE, M.-J. & McGEOCH, D. J. (1982). DNA sequence analysis of an immediate-early gene region of the herpes simplex virus type 1 genome (map coordinates 0.950 to 0.978). *Journal of General Virology* **62**, 1-15.
- OAKES, J. E., ILTIS, J. P., HYMAN, R. W. & RAPP, F. (1977). Analysis by restriction enzyme cleavage of human varicella-zoster virus DNAs. *Virology* **82**, 353-361.
- OSTROVE, J. M., REINHOLD, W., FAN, C.-H., ZORN, S., HAY, J. & STRAUS, S. E. (1985). Transcription mapping of the varicella-zoster virus genome. *Journal of Virology* **56**, 600-606.
- PRESTON, V. G. & FISHER, F. B. (1984). Identification of the herpes simplex virus type 1 gene encoding the dUTPase. *Virology* **138**, 58-68.
- PUSTELL, J. & KAFATOS, F. C. (1982). A high speed, high capacity homology matrix: zooming through SV40 and polyoma. *Nucleic Acids Research* **10**, 4765-4782.
- QUINN, J. P. & McGEOCH, D. J. (1985). DNA sequence of the region in the genome of herpes simplex virus type 1 containing the genes for DNA polymerase and the major DNA binding protein. *Nucleic Acids Research* **13**, 8143-8163.
- RICHARDS, J. C., HYMAN, R. W. & RAPP, F. (1979). Analysis of the DNAs from seven varicella-zoster virus isolates. *Journal of Virology* **32**, 812-821.
- RIXON, F. J. & CLEMENTS, J. B. (1982). Detailed structural analysis of two spliced HSV-1 immediate-early mRNAs. *Nucleic Acids Research* **10**, 2241-2256.
- RIXON, F. J. & McGEOCH, D. J. (1984). A 3'-coterminal family of mRNAs from the herpes simplex virus type 1 short region: two overlapping reading frames encode unrelated polypeptides one of which has a highly reiterated amino acid sequence. *Nucleic Acids Research* **12**, 2473-2487.
- ROBERTS, T. M., SWANBERG, S. L., POTEETE, A., RIEDEL, G. & BACKMAN, K. (1980). A plasmid cloning vehicle allowing a positive selection for inserted fragments. *Gene* **12**, 123-127.
- SANGER, F., COULSON, A. R., BARRELL, B. G., SMITH, A. J. H. & ROE, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**, 161-178.
- STADEN, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**, 4731-4751.
- STADEN, R. & MC LAUCHLAN, A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Research* **10**, 141-156.
- STOW, N. D. & DAVISON, A. J. (1986). Identification of a varicella-zoster virus origin of DNA replication and its activation by herpes simplex virus type 1 gene products. *Journal of General Virology* **67**, 1613-1623.
- STRAS, S. E., AULAKH, H. S., RUYECHAN, W. T., HAY, J., CASEY, T. A., VANDE WOUDE, G. F., OWENS, J. & SMITH, H. A. (1981). Structure of varicella-zoster virus DNA. *Journal of Virology* **40**, 516-525.
- STRAS, S. E., OWENS, J., RUYECHAN, W. T., TAKIFF, H. E., CASEY, T. A., VANDE WOUDE, G. F. & HAY, J. (1982). Molecular cloning and physical mapping of varicella-zoster virus DNA. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 993-997.
- STRAS, S. E., HAY, J., SMITH, H. & OWENS, J. (1983). Genome differences among varicella-zoster virus isolates. *Journal of General Virology* **64**, 1031-1041.
- TAMURA, T., NODA, M. & TAKANO, T. (1981). Structure of the baboon endogenous virus genome: nucleotide sequences of the long terminal repeat. *Nucleic Acids Research* **9**, 6615-6626.
- TAYLOR, P. (1984). A fast homology program for aligning biological sequences. *Nucleic Acids Research* **12**, 447-455.
- TAYLOR, P. (1986). A computer program for translating DNA sequences into protein. *Nucleic Acids Research* **14**, 437-441.
- TOMITA, M. & MARCHESI, V. T. (1975). Amino acid sequence and oligosaccharide attachment sites of human erythrocyte glycoporphin. *Proceedings of the National Academy of Sciences, U.S.A.* **72**, 2964-2968.
- VIEIRA, J. & MESSING, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259-268.
- WAGNER, E. K. (1985). Individual HSV transcripts. Characterization of specific genes. In *The Herpesviruses*, vol. 3, pp. 45-104. Edited by B. Roizman. New York: Plenum Press.

- WATSON, R. J. & VANDE WOUDE, G. F. (1982). DNA sequence of an immediate-early gene (IE mRNA-5) of herpes simplex virus type 1. *Nucleic Acids Research* **10**, 979-991.
- WATSON, R. J., SULLIVAN, M. & VANDE WOUDE, G. F. (1981a). Structures of two spliced herpes simplex virus type I immediate-early mRNAs which map at the junctions of the unique and reiterated regions of the virus DNAs component. *Journal of Virology* **37**, 431-444.
- WATSON, R. J., UMENE, K. & ENQUIST, L. W. (1981b). Reiterated sequences within the intron of an immediate-early gene of herpes simplex virus type 1. *Nucleic Acids Research* **9**, 4189-4199.
- WELLER, S. K., SPADARO, A., SCHAFER, J. E., MURRAY, A. W., MAXAM, A. M. & SCHAFER, P. A. (1985). Cloning, sequencing, and functional analysis of ori_L, a herpes simplex virus type I origin of DNA synthesis. *Molecular and Cellular Biology* **5**, 930-942.
- ZWEERINK, H. J., MORTON, D. H., STANTON, L. W. & NEFF, B. J. (1981). Restriction endonuclease analysis of the DNA from varicella-zoster virus: stability of the DNA after passage *in vitro*. *Journal of General Virology* **55**, 207-211.

(Received 7 April 1986)